VIETNAM ACADEMY OF SOCIAL SCIENCES

# Market, Policy and Poverty Reduction in Vietnam



VIETNAM CULTURE AND INFORMATION PUBLISHING HOUSE

Vietnam Academy of Social Sciences

# Market, Policy and Poverty Reduction in Vietnam

*Editors:* HENRIK HANSEN University of Copenhagen

NGUYEN THANG Vietnam Acedamy of Social Sciences

# Table of Contents

Forew	ord	i		
Acknowledgements				
Contri	butors	iv		
List of	Abbreviations	v		
Summ	ary of the Volume	vi		
PART	I: THE WORKING OF THE MARKET AND POVERTY REDUCTION	1		
Chapt	er 1. Testing for Food market integration:	3		
	A study of the Vietnamese paddy market			
1.	Introduction	4		
2.	Paddy Production, Distribution and Marketing	5		
3.	Modelling and Testing for Market Integration with Transfer Costs	13		
4.	Data on Paddy Prices, Transfer Costs and Time Series Properties	17		
5.	Lests for Market Integration	23		
6.	Conclusion and Policy Implications	28		
	Keterences	29		
Chapt	er 2. Regional Labor Market Integration in Vietnam	30		
1.	Introduction	30		
2.	Macro versus micro convergence	31		
3.	Explaining regional wage differentials	37		
4.	Labor market integration under market imperfections	42		
5.	Conclusions	50		
	References	50		
Chapt	er 3. Determinants and Impacts of Migration in Vietnam	59		
1.	Introduction	59		
2.	Migration Patterns in Vietnam	62		
3.	Determinants of migration in Vietnam	68		
4.	Migration, remittances and income inequality	77		
5.	Discussion and conclusions	88		
	References	89		
PART	II: POLICY AND POVERTY REDUCTION	93		
Chapt	er 4. Better Than its Reputation? The Incidence of Social Transfers and Education Foo Exampliancin Vietnam	95		
1	Introduction	OF		
1. ว	Introduction A Dublic Expanditum View on Social Transfers and Education For Examplian	95		
2.	A Fublic Expenditure view on Social transfers and Education Fee Exemption	99 105		
J.	The incidence of again transfers and education fee exemptions	105		
4.	I ne incluence of social transfers and education fee exemptions in 2004	112		
5.	Social Transfers and Poverty 2002-2004	123		
6.	Conclusion and policy implementations	125		
	Kelerences	126		

Chapt	er 5. Ethnicity and Poverty Reduction	134
1.	Introduction	134
2.	Government Policies towards Ethnic Minorities	135
3.	The Socio-Economic Development of the Ethnic Groups	137
4.	Econometric Analysis of the Majority/Minority Expenditure Gap	147
5.	Summary and Concluding Remarks	162
	References	164
Chapt	er 6. Changes in Household's Poverty Status in Rural Vietnam 2002-2004	169
1.	Introduction	169
2.	Poverty Transitions from 2002 to 2004	170
3.	An Econometric Analysis of the Poverty Transitions	180
4.	Conclusion	194
	References	195
Chapt	er 7. A Look at Parameter Constancy and	197
	Poverty Predictions in Poverty Mapping Models	
1.	Introduction	197
2.	The estimation technique and prediction comparison statistics	200
3.	The data	205
4.	Comparison of regression results for three household surveys	209
5.	Predictions of provincial poverty rates based on the Census 1999	215
6.	Conclusion	222
	References	224
Annex	: Vietnam Household Living Standards Survey (VHLSS),	228
	2002 and 2004 - Basic Information	
1.	Introduction	228
2.	The Survey Questionnaires	229
3.	The Sample	242
4.	Implementation	252
5.	Data Processing	254
6.	Dissemination Policy	254

### Foreword

During the 1990s and the first half of the first decade of the 21<sup>st</sup> century, Vietnam achieved rapid and broad-based economic growth, resulting in remarkable progress in reducing poverty in its multiple dimensions. Poverty assessments conducted in 1992-93, 1998-1999, 2002-2003 and 2004 have revealed that Vietnam is one of the few countries in the world in which poverty, however measured, declined dramatically within the relatively short period of 1993-2004. During this period, along with the growth of real GDP per capita by 5.9 percent per year on average, the share of poor people in this country dropped by two thirds and approximately 30 million people were lifted out of poverty.

Vietnam, however, is likely to encounter numerous challenges in sustaining a similarly rapid pace of poverty reduction. First, the poverty reducing effect of economic growth may become smaller than it has been until very recently. In other words, a higher growth rate is required to reduce each percentage point of the poverty rate while, in turn, each percentage point of economic growth may require a higher level of investment (or alternatively, forgone consumption) if Viet Nam cannot improve the quality and efficiency of investment. This implies that poverty reduction may become more costly in the years to come. Second, the lack of progress in poverty reduction among ethnic minorities, who represent approximately 14 percent of the population, but account for some 39 percent of the poor in 2004, is a serious cause for concern in Vietnam. Third, although relative inequality as measured by the Gini coefficient rose only modestly from 0.33 in 1993 to 0.37 in 2004, the absolute gap between the rich and the poor, and between rural and urban dwellers persists, with poverty becoming largely a rural phenomenon. There are also considerable regional disparities in poverty rates and the pace of poverty reduction. The northern mountains, the north central coast and the central highland, which together make up 57 percent of the total poor population of Vietnam, all have a high poverty rate of over 30 percent in 2004, as opposed to the national average of under 20 percent. Fourth, the relatively large number of non-poor people whose consumption expenditure is just above the poverty line level indicates that protecting vulnerable people from falling back into poverty is becoming increasingly important for the reduction in overall poverty, particularly in light of various shocks they face now and in the future as Vietnam intensifies the international integration process. Finally, as Vietnam is expected to graduate from the list of poorest countries by 2010, it is likely to be confronted with new and increasingly sophisticated challenges in the coming years.

In this context, maintaining the rapid and pro-poor pattern of economic growth will require, among other things, a gradual shift towards evidence-based policy making, whereby policies are grounded in research evidence generated using sound methodologies and reliable data. The project, "Support for Strengthening Evidence-based Pro-Poor Policy Making in Vietnam", which was carried out by the Vietnam Academy of Social Sciences (VASS), was an effort to enhance the analytical skills of young Vietnamese researchers in doing policy-relevant and poverty-focused research. Under this project, a number of empirical studies were carried out with a strong emphasis on the use of available micro level datasets such as VLSS93, VLSS98, VHLSS02, VHLSS04, the GSO's enterprise surveys and censuses, among others. Young and capable researchers coming from a number of Vietnamese

research institutions and universities carried out these studies under the direct guidance of international experts acting as their academic mentors. This approach, with a strong emphasis on "learning by doing", aimed to help the participating young researchers to better prepare themselves for a much greater role in the Vietnamese research community in the near future.

The studies carried out under this project have resulted in a number of papers; each of them is presented as one chapter of this book. The issues addressed in the book can be classified into two broad groups, which are presented in Part I and Part II respectively. Part I, which consists of the first three chapters, analyses various aspects of the working of the goods and labor markets and their implications for poverty reduction in Vietnam. Part II examines some important policies from the perspective of their contribution to poverty reduction. The final chapter evaluates the use of poverty mapping models for poverty prediction, which has become quite popular in recent years. The annex of the book provides technical details on the household surveys VHLSS02 and VHLSS04, which were extensively used in the studies presented in this book. Though being far from comprehensive, the book will undoubtedly be of interest to the research community in and outside Vietnam. Selected research findings were used as important inputs for the Vietnam Poverty Update Report 2006, which was published by VASS and presented and circulated to a wide range of policy makers at national and sub-national levels in Vietnam, and revised versions of some of the chapters will appear in international economic journals.

Prof. DO HOAI NAM President Vietnam Academy of Social Sciences

### Acknowledgments

This book presents findings of studies conducted under the project, "Support for Strengthening Evidence-based Pro-Poor Policy Making in Vietnam", which was carried out over the period from December 2004 to June 2006 by the Vietnam Academy of Social Sciences (VASS) with financial support from the World Bank managed ASEM trust fund. During the process of project implementation, VASS also received financial and in-kind support from the Royal Danish Embassy and DANIDA, which allowed for a resident senior academic advisor, Prof. Henrik Hansen of the University of Copenhagen. Prof. Hansen is the lead editor of this volume, as well as a mentor to younger researchers and a co-author of a number of the book's chapters.

The contributions of the authors of each chapter, whose names are listed on the next page, are duly acknowledged. During the research process and the preparation of the book, the editors and chapter authors received valuable support from a number of people, including Nguyen Thi Thanh Ha, Nguyen Thu Hang, Pham Anh Tuyet, Do Thu Huong and Nguyen Thi Hai Oanh (VASS), and Nguyen Thi Song Anh, Dao Kim Lan and Phan Bich Khanh (consultants). We would also like to extend our sincere thanks to Prof. Do Hoai Nam, President of VASS, Dr. Martin Rama, Lead Economist, and Mr. Dinh Tuan Viet – Economist at the World Bank in Vietnam for the encouragement and support they provided throughout the project cycle. There are also numerous other people who contributed in one way or another. Although we will not attempt to list them all, their assistance is highly appreciated.

### *Contributors*

Bob Baulch	Institute of Development Studies, University of Sussex, Brighton, United Kingdom
Henrik Hansen	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam; Department of Economics, University of Copenhagen, Denmark
Hoàng Thanh Hà	Central Institute for Economic Management, Vietnam
Hoàng Thanh Hương	Hanoi National Economics University, Vietnam
Lê Đặng Trung	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam
Lê Văn Chơn	Danang University, Vietnam
Nguyễn Phong	General Statistical Office
Nguyễn Thắng	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam
Nguyễn Thị Nguyệt	Central Institute for Economic Management, Vietnam
Nguyễn Thu Phương	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam
Nguyễn Việt Đức	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam
Phạm Hương Giang	Institute of Agricultural Economics, Vietnam
Phạm Thùy Giang	Vietnam Chamber of Commerce and Industry, Vietnam
Phùng Đức Tùng	General Statistical Office
Remco Oostendorp	Free University Amsterdam, the Netherlands; Tinbergen Institute, the Netherlands
Trần Bình Minh	Central Institute for Economic Management, Vietnam
Trần Ngô Thị Minh Tâm	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam
Trương Thu Trang	Institute of Agricultural Economics, Vietnam
Vũ Hoàng Đạt	Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, Vietnam

## List of abbreviations

ADB	Asian Development Bank
ADF	Augmented Dickey-Fuller
CIEM	Central Institute for Economic Management
ERS	Elliott, Rothenberg, and Stock's (1996)
FDI	Foreign Direct Investment
GDP	Gross Domestic Product
GMM	General Methods of Moments
GSO	General Statistics Office
HCMC	Hochiminh City
HEPR	Hunger Eradication and Poverty Reduction
ICARD	Information Center for Agriculture and Rural Development
ILSSA	Institute of Labor Sciences and Social Affairs
MARD	Ministry of Agriculture and Rural Development
MOET	Ministry of Education and Training
MOLISA	Ministry for Labor, Invalids and Social Affairs
MRD	Mekong River Delta
NA	Not available
NTPs	National Targeted Programs
RRD	Red River Delta
SGFs	social guarantee funds
SOEs	Stated-Owned Enterprises
SRV	Socialist Republic of Vietnam
TV	Television
TSLS	Two-stage Least Square
UNDP	United Nations Development Programme
UNIDO	United Nations Industrial Development Organization
VASS	Vietnam Academy of Social Sciences
VHLSS	Vietnam Household Living Standard Survey
VLSS	Vietnam Living Standard Survey
VND	Vietnam Dong
VSI	Vietnam Social Insurance

### Summary of the Volume

This book consists of seven chapters, which are grouped in two parts. The first part is focused on analysis of some important aspects of the working of the goods and labor markets and their implications for poverty reduction in Vietnam. The second part examines some important policies from the perspective of their contribution to poverty reduction. The book also includes one annex, which provides technical details on the household surveys VHLSS 02 and VHLSS 04.

The first chapter, "Testing for Food Market Integration: A Study of the Vietnamese Paddy Market", written by Le Dang Trung, Tran Ngo Minh Tam, Bob Baulch and Henrik Hansen, looks into the issue of spatial integration of the national paddy market, which arguably plays a crucial role in the livelihoods of many farmers in Vietnam. With its increasing integration into the global economy, agricultural exports and rural incomes in Vietnam have increased substantially in recent years. At the sub–national level, however, there are still concerns that not all regions and categories of agricultural producers have and will benefit from the ongoing liberalization of agricultural markets, leading to widening regional gaps. Vietnam's elongated geography and lack of spatial market integration pose special problems in this regard.

This paper attempts to answer three interrelated questions: (a) whether there is spatial integration between paddy markets in the North and South of Vietnam; (b) whether there is spatial integration in paddy markets within the North and within the South of Vietnam; and, (c) if intra-regional integration is stronger and faster than inter-regional integration.

The empirical model that the authors develop to answer these questions uses newly gathered data on transfer costs to generalize the well known model of spatial market integration, due to Ravallion to allow for the possibility of threshold effects. A sequential testing strategy is developed which progressively tests for market segmentation, the number of thresholds, long-run market integration, common dynamics/informational efficiency, and (a strict version of) the 'Law of One Price' within an error-correction framework.

When the unrestricted version of this model is estimated using monthly paddy prices for eight markets between 1993 and 2006, the authors find weak evidence of market integration between paddy markets in the North and South of Vietnam with an absence of threshold effects. However, there is evidence of both threshold effects and stronger forms of spatial market integration for paddy markets within the North and within the South, with at least 60 percent of price changes being transmitted between markets (within one month) whenever price spreads exceed their upper or lower thresholds. The extent and speed of price transmission within regional paddy markets is generally faster in the South than in the North of Vietnam. However, the instantaneous version of the 'Law of One Price', which requires full price adjustment to occur within one month, only holds for a few regimes and market pairs.

Three main policy implications flow from these findings. First, since there is limited evidence of integration between paddy markets in the North and South of Vietnam, national level policies cannot be relied upon to stabilize or support paddy prices. Second, since there is evidence of spatial market integration within the Red River and Mekong River deltas, paddy markets within these regions can be

relied upon to transmit price signals between deficit and surplus areas relatively well. Third, since the speed and extent of price transmission is relatively rapid within the North and within the South of Vietnam, the private sector trade can be relied upon to transfer rice and paddy between markets in an efficient manner. Problems might, however, emerge if large demand-supply imbalances were to emerge between the North and South, as transfer costs would prevent private sector trade taking place. In these circumstances, the public sector might need to intervene, in a consistent and market friendly way, to ensure adequate food supplies in the short-term.

The next two chapters analyze various aspects of the labor market in Vietnam. Chapter 2, "Regional Labor Market Integration" is written by Hoang Thanh Huong, Le Van Chon, Le Dang Trung and Remco Oostendorp. Using evidence from four waves of household surveys over the period 1993-2004, the paper finds that regional wage levels on the whole diverged during the period 1993-2004, although there is evidence of convergence more recently (2002-2004). Regional wage gaps can increasingly be explained by regional differences in human capital, industrial structures and ownership. Even after controlling for (changes in) these factors, however, hourly wages are still diverging across regions in Vietnam. The authors do, however, find that regional wage gaps with *neighbouring* regions are converging after they allow for the regional differences in human capital, industrial structures and ownership. Combining these findings tends to suggest that labor markets in Vietnam are *locally converging but globally diverging*.

This chapter also estimates regional shadow wages in agricultural self-employment. It finds that shadow wages are significantly lower than market wages, confirming a lack of integration between wage- and self-employment and the existence of surplus labor in rural areas. However, shadow wages as a ratio of market wages increased between 2002 and 2004 for the whole country, suggesting that the markets for wage- and self-employment are becoming more integrated in Vietnam more recently. The degree of integration varies across regions; it is found to be the weakest in the North East and the North West, and the strongest in the South East and the Mekong River Delta.

In summary, for Vietnam as a whole, integration of regional labor markets has been generally weak. There are, however, signs of increasing local integration.

Chapter 3, "Determinants and Impacts of Migration in Vietnam", prepared by Nguyen Thu Phuong, Tran Ngo Minh Tam, Nguyen Thi Nguyet and Remco Oostendorp uses the recent Vietnam Household Living Standard Survey 2004 to make its analyses. Most of the previous studies on the determinants and impacts of migration have focused on destination rather than origin areas of migration. This limits our understanding of the determinants of migration and also does not provide evidence on important impacts of migration such as on household inequality in origin areas.

In terms of determinants of migration, the study shows that migration is a highly selective process, which is strongly affected by household and commune characteristics, although differently according to the type of migration and across urban and rural areas. The authors find evidence of the existence of a 'migration hump' for economic long-term migration, with an inverted U-shape in the probability of migration with respect to per capita expenditures. The presence of non-farm employment opportunities reduces short-term migration but not long-term out-migration for economic reasons.

In terms of impacts, the study analyses the impact of migration on household expenditures and household inequality. Migration is found to have a strong positive impact on household expenditures but, at the same time, increases the Gini coefficient of per capita household expenditures from 0.38 to 0.42 in origin areas compared to the non-migration case.

Part II of this book starts with Chapter 4, "Better Than its Reputation? The Incidence of Social Transfers and Education Fee Exemptions in Vietnam", which is written by Henrik Hansen and Le Dang Trung. This chapter analyzes the contribution of budget transfers to poverty reduction in Vietnam. Together with high economic growth over the last few years, redistribution of public funds has been a crucial topic for both researchers and policy makers. Since 1997 there have been at least six

studies on the incidence of public transfer programs which make use of the 1992/93 and 1997/98 VLSSs. These analyses do not recognize the success of the Vietnamese government's poverty-targeted redistributive schemes as they conclude that the social transfer system is regressive, and that it does not provide effective social protection to the majority of the poor. With the availability of new data, namely, the 2002 and 2004 household surveys (VHLSS02 and VHLSS04), the incidence analysis can be updated. The authors seek answers to the following questions: what is the incidence of the redistribution programs? Are they targeted to the poor? What are the roles of these programs in poverty dynamics? In doing so, the authors challenge many of the previous studies. They raise concerns about: (i), the choice of the counter-factual welfare measure in existing studies and (ii) the suitability of the household survey data for conducting the analyses.

In estimating the counter-factual welfare measure, which is based on an estimate of the marginal propensity to consume out of transfer (*mpct*), the authors notice that the *mpct* seems not to be precisely estimated. Therefore, even though the point estimate derived is one, the authors cannot reject the hypothesis that the true value is only one-half of that found in an earlier study. In fact, they come to conclude that the *mpct* could lie anywhere from about 0.1 to more than 1. Thus, the social transfer incidence results should be interpreted with great caution.

In the incidence analysis, the authors look at two types of transfers and two types of education fee exemptions: social insurance and social assistance, and tuition fee exemptions and school contribution fee exemptions, respectively. The results paint a somewhat different picture of the distribution of social transfers than the studies based on the 1992/93 and 1997/98 household surveys. First of all, social transfers were not regressive in 2004. Second, the transfers had a sizeable impact on the level of poverty in both 2002 and 2004 and, furthermore, provided protection against falling into poverty in that period.

Chapter 5, "Ethnicity and Poverty Reduction", written by Hoang Thi Thanh Huong, Pham Huong Giang, Tran Binh Minh and Henrik Hansen seeks to identify factors that explain the lack of progress in poverty reduction among ethnic minorities against the backdrop of rapid overall poverty reduction. Large differences in poverty rates and other socio-economic measures between the major ethnic groups (the Kinh people often grouped with the Hoa people) and the other 52 ethnic minority groups living in Vietnam is a well established fact, and is already one of the biggest policy concerns in Vietnam at present.

This chapter updates the quantitative analyses already conducted in previous studies to include the latest household survey information (VHLSS 2004). The chapter also seeks to illustrate changes over time in the relative importance of household endowments, the returns to household endowments and household location. In previous studies these were found to be three inter-linked sources of the large, and increasing, discrepancy in majority and minority well-being during the 1990s. The chapter does this by using directly comparable empirical models for per capita consumption expenditure and comparable samples in an econometric analysis of the expenditure gaps between majority and minority households.

Using household level data from 1993, 1998 and 2004 the authors find the decline in poverty among the ethnic minorities to be markedly larger in the second half of the period than in the first half. It is also found that although all ethnic groups have experienced improvements in well-being over the last 10-15 years, the majority people have done far better than the minority people.

Household demographic characteristics and assets differ between the majority and minority people. There are also discernible differences in a household's key assets between the different groups, such as the formal education level of the head of the household, non-farm employment, and household landholdings. These household characteristics and endowments are included as regressors in an econometric analysis of the gap in per capita real expenditure conditioned on location effects by including commune fixed effects. The paper also seeks to 'remove' the location effects by analyzing

sub-samples of communes in which the majority and minority are both present (mixed communes). The regression results show somewhat surprising similarities in the expenditure models, in contrast to earlier analyses of the expenditure gap done by van de Walle and Gunewardena (2001), and Baulch *et al.* (2004). In particular, the authors do not find significant difference in the returns to education or in the returns to non-farm employment.

The regression models for the Kinh-Hoa majority and the minority groups are used to decompose the average per capita real expenditure gap into factors relating to household compositions and endowments (household characteristics), and factors relating to returns to the household characteristics. The authors emphasize that the results of the decompositions should be interpreted with care due to the inclusion of commune fixed effects. Location is part of the household endowment and, as such, not all of it should be attributed to differences in returns. Hence, the importance of returns is overestimated while the importance of endowments is underestimated if the decomposition is interpreted in a strict sense.

The decomposition analysis shows that the relative contribution of household characteristics and returns, plus location effects, is surprisingly constant, as household characteristics only account for about 20 percent of the expenditure gap in all three years. When the location effect is removed, by analyzing the sub-sample of mixed communes, the results change somewhat as the gap then decreases from 1998 to 2004. But it is still significant in 2004. The most interesting result of the mixed commune analysis is that in 2004 the expenditure gap is probably only caused by differences in household endowments. Overall, the decompositions point to location as having a strong impact on the average expenditure gap, supporting the view that location is an important obstacle to increasing the well being of minority people. A pure, 'ethnicity factor', however, cannot be ruled out since both the descriptive analysis and the econometric results clearly illustrate differences in household endowments in communes in which Kinh-Hoa and minority people live together.

A number of policy implications have been drawn on the basis of these research findings. First, the strong effect of location on ethnic well being calls for continued and increased government efforts in extremely difficult ethnic and mountainous area. Targeted programs such as Program 135, which was initiated in 1998, has been found to be well targeted but the coverage of the program is probably too limited. Hence, a scaling up of the most successful parts of that program is probably a good idea. Second, the presence of the location effect implies that it is important to 'localize' government support programs by giving consideration to the local environment in the design of specific programs. This calls for in-depth analyses of income and poverty at the regional level and possibly even at lower geographical levels. For example, a rural development strategy to increase the labor and land productivity of crop production would help increase the return to the only assets that seems to be in relative abundance among the ethnic minority people, and therefore it would be likely to benefit the ethnic minority groups in remote areas.

Chapter 6, "Changes in Household's Poverty Status in Rural Vietnam 2002-2004", which is prepared by Vu Hoang Dat, Truong Thi Thu Trang, Nguyen Viet Duc and Henrik Hansen seeks to establish an 'extended' poverty profile for Vietnam that not only characterizes the poor and non-poor at a given point in time but, rather, describes movements in and out of poverty. The authors use the panel data component of the rural parts of the two household surveys VHLSS 2002 and VHLSS 2004. The panel data shows a large outflow from poverty, about 15 percent, in the period and a relatively smaller inflow, of less than 5 percent, to poverty. Still, a little less than 20 percent of the rural population was poor in both 2002 and 2004.

In the study of the gross flows in and out of poverty the variations in the flows are described using location, education, occupation, agricultural production, household assets, local infrastructure and ethnicity as the main covariates. This is very similar to the choice of covariates in more standard poverty profile studies. The variations in the gross flows are described using both cross tabulations

and an econometric model (a multinomial logit model) that analyzes the joint impact of the covariates. The results of the analyses are clearly in line with the more standard poverty profiles, but they illustrate the importance of some of the poverty determinants in a dynamic perspective. In particular, the study shows that the ethnic minorities in Vietnam have a much smaller probability of escaping poverty than the ethnic majority consisting of the Kinh and the Hoa (Chinese) and this is so even when differences in location, family structure, education and occupation are taken into account. Among the other determinants the authors find that both education and non-farm jobs are reducing the probability of becoming and being poor. Primary schooling reduces the risk of falling into poverty whilst secondary schooling both reduces the risk of falling into poverty and increases the chance of escaping poverty. A clear policy indication of this finding is that the Government of Vietnam should continue its effort in encouraging people living in the rural areas to engage in and complete schooling at the secondary level. For non-farm occupations the analysis shows that some occupations such as production and sales have protected against falling into poverty while a larger fraction of white-collar workers have escaped poverty, which is interesting as the poverty rate among white-collar workers is already very low in 2002.

Finally, the authors stress that non-farm work is not the only way out of poverty in the rural parts of Vietnam. A larger agricultural production, or higher prices on agricultural commodities are also associated with a better chance of escaping poverty. Hence, agricultural policies targeted at increasing the productivity in agriculture and credit policies facilitating increased livestock production will in all likelihood continue to be poverty reducing in the future.

The final chapter, "A Look at Parameter Constancy and Poverty Predictions in Poverty Mapping Models", by Henrik Hansen, Pham Huong Giang and Vu Hoang Dat, evaluates the validity of using poverty mapping models for poverty prediction. This is a very useful exercise, as mappings have become an increasingly popular tool in investigations of social, environmental and economic problems in many developing countries; currently about 25 developing countries have produced detailed poverty maps to assist local and central policy makers.

Vietnam has poverty maps with estimates of the poverty incidence at the commune level (Minot, Baulch and Epprecht, 2003) and at the provincial level (Minot and Baulch, 2004). These maps are estimated by combining the Vietnamese living standard measurement survey from 1998 with the 1999 Population and Housing Census. Because of the rapid socio-economic development in Vietnam, there is a growing interest in updating these poverty maps to more recent years. Even though there are new living standard measurement surveys (VHLSS 2002 and VHLSS 2004), a new census will not be available until the end of the decade. Therefore, in order to update the poverty maps solutions other than waiting for a new census need to be investigated. One possibility is to update the poverty maps by using a panel component of two surveys. This option is possible in Vietnam as the VHLSS 2002 and VHLSS 2004 have a panel component of about 4,000 households. A first step in such a panel based update of the poverty map is to assess if the VHLSS 2002 in combination with the Population and Housing Census from 1999 provides a good estimate of the local area poverty incidences.

This study makes a first pass by comparing provincial level poverty incidences, which are based on consumption models using the three surveys VLSS 1998, VHLSS 2002 and VHLSS 2004. The authors first compare the population estimates of the regressors in the consumption function across the three surveys and the census. Surprisingly, the result is that, in general, none of the three surveys match the census based estimated population means. Hence, it is not clear that any of the three surveys represents the same underlying populations as the census. While this is somewhat unexpected considering that the time spread between the VLSS 1998 and the census is only one year, it needs not invalidate the poverty mapping results. But, the result implies that there are no strong statistical reasons for preferring the VLSS 1998 in the poverty mapping to the VHLSS 2002. In contrast, one may

argue in favor of the VHLSS 2002 because of the much larger sample size that allows for a more detailed consumption model.

The second test of the poverty mapping models is a test of parameter constancy across consumption models, using data from each of the three household surveys. The study finds that the consumption model parameters are statistically different when they are estimated using different surveys. This means that the systematic components captured in the consumption model are changing over time. This may be interpreted as a clear rejection of a constant underlying super-population.

Because the poverty incidence predictions are non-linear functions of the consumption function parameters, non-constancy of the latter parameters does not imply that the poverty predictions will be statistically different. Therefore, the authors use each of the three surveys in combination with the census to estimate provincial level poverty incidences. Subsequently, they test if the predicted poverty incidences are statistically different. It turns out that the predictions based on the combination of the VLSS 1998 and the census, are not statistically different from the predictions generated from the combination of the VHLSS 2002 and the census. In contrast, it is found that the last combination, the VHLSS 2004 and the census, results in statistically different predictions. Based on this statistical analysis the authors conclude that the combination of the VHLSS 2002 and the census generates poverty incidence predictions that appear to be valid estimates of the poverty situation in 1999.

The conclusion opens up for further poverty mapping analyses using the VHLSS 2002. First of all, it would be interesting if the same conclusion is reached at the commune level, *i.e.*, by combining the survey with the 33 percent sample of the Population and Housing Census 1999. Second, it allows for more flexibility in the consumption models because the sample in the VHLSS 2002 is very large. In particular, it seems worthwhile to analyze if allowing for parameter variation across ethnic majority/minority groups can improve the predictions. Furthermore, stratum level estimations of the consumption function parameters can be estimated with greater precision, and this may also result in improved predictions. Finally, the poverty map can possibly be updated by combining the data in the VHLSS 2002 and the VHLSS 2004 as discussed above.

# PART I: THE WORKING OF THE MARKET AND POVERTY REDUCTION

# 1

### Testing for Food Market Integration: A Study of the Vietnamese Paddy Market

#### Le Dang Trung, Tran Ngo Thi Minh Tam, Bob Baulch and Henrik Hansen

With its increasing integration into the world economy, agricultural exports and rural incomes in Vietnam have increased substantially in recent years. At the sub–national level, however, there are concerns that not all regions and categories of agricultural producers have and will benefit from the ongoing liberalization of agricultural markets.

Vietnam's elongated geography and lack of spatial market integration pose special problems in this regard. Accordingly, this study aims to answer three interrelated questions: (a) whether there is spatial integration between paddy markets in the North and South of Vietnam; (b) whether there is spatial integration in paddy markets within the North and within the South; and, (c) if within-region integration is stronger and faster than between-region integration.

The empirical model we develop to answer these questions, uses estimates of transfer costs to generalize the well known model of spatial market integration due to Ravallion to allow for the possibility of threshold effects. A sequential testing strategy is developed which progressively tests for market segmentation, the number of thresholds, long-run market integration, common dynamics/informational efficiency, and (a strict version of) the 'Law' of One Price within an error-correction framework.

When the unrestricted version of this model is estimated using monthly paddy prices for eight markets between 1993 and 2006, we find weak evidence of market integration between paddy markets in the North and South of Vietnam with an absence of threshold effects. However, there is evidence of both threshold effects and stronger forms of spatial market integration for paddy markets within the North and within the South, with at least 60% percent of price changes being transmitted between markets within one month whenever price spreads exceeds their upper or lower thresholds. The extent and speed of price transmission within regional paddy markets is generally faster in the South than the North of Vietnam. However, the instantaneous version of the 'Law' of One Price, which requires full price adjustment to occur within a month, only holds for a few regimes and market pairs.

Three main policy implications flow from these results. First, since there is limited evidence of integration between paddy markets in the North and South of Vietnam, national level policies cannot be relied upon to stabilize or support paddy prices. Second, since there is evidence of spatial market integration within the Red River and Mekong River deltas, paddy markets within these regions can be relied upon to transmit price signals between deficit and surplus areas relatively well. Third, since the

speed and extent of price transmission is relatively rapid within the North and within the South of Vietnam, the private sector trade can be relied upon to transfer rice and paddy between markets in an efficient manner. Problems might, however, emerge if large demand-supply imbalances were to emerge between the North and South, as transfer costs would prevent private sector trade taking place. In these circumstances, the public sector might need to intervene, in a consistent and market friendly way, to ensure adequate food supplies in the short-term.

#### 1. Introduction

With its increasing integration into the world economy, Vietnam's agricultural exports have grown substantially in recent years. From being a net importer of rice in the late 1980s, by the late 1990s Vietnam had become one of the largest rice exporting countries. Paddy production has grown from about 23 million metric tons in 1993 to around 36 million metric tons in 2005, with 70 percent of total paddy output coming from the Red River and Mekong River deltas.

At the sub-national level, there are concerns that not all regions and categories of agricultural producers have and will benefit from the liberalization of Vietnam's agricultural markets. There are two distinct aspects to these concerns. First, if domestic markets are not spatially integrated, not all regions will benefit from market and trade liberalization to the same extent. The elongated geography of Vietnam poses special problems in this regard. Second, even if domestic markets are integrated in the long term, lack of integration in the short-term may mean that price changes are not transmitted between consumption centres and production areas quickly enough. If this is the case, agricultural market liberalization may disproportionately benefit traders and agricultural processors rather than agricultural producers and consumers.

Accordingly, this study aims to answer three interrelated questions. The first question is whether there is spatial integration between paddy markets in the North and South of Vietnam. The elongated geography of Vietnam results in high transfer costs for food market transactions between the northern and the southern regions. This makes it interesting to look into the degree of integration in the paddy markets, as trade is in all likelihood not the determining factor for market integration between the North and the South of Vietnam.

The second question we seek to answer is whether there is spatial integration in paddy markets within the North and within the South, specifically in the Red River and Mekong River deltas. We focus on paddy prices in the Red River and Mekong deltas for this analysis, both because they produce over two-thirds of national paddy output and because the price data for these regions is fairly comprehensive and complete.

Finally, a natural third question to ask is if within-region integration is stronger and faster than between-region integration.

Our analysis, which uses threshold error correction models to sequentially test for different forms of market integration using monthly paddy prices for eight provinces in Vietnam, indicates a weak integration of the paddy markets between the North and the South of Vietnam. However, within the northern and southern regions, there is evidence of stronger market integration, with 50 to 85 percent of price changes being transmitted between markets within one month whenever inter-market price spreads are large enough to induce profitable trade flows between provinces.

This chapter is structured as follows. In Section 2 we provide an overall description of the paddy production in Vietnam from the beginning of the 1990s to 2005. Section 3 discusses the importance of thresholds when analyzing spatial market integration in the presence of transfer costs and describes our modelling and testing strategy. In Section 4 we describe the paddy price data used and how we have estimated transfer costs. Descriptive statistics and (linear) tests of stationarity and cointegration

are also presented in this section. In Section 5, we use the paddy price series to test for various forms of market integration between the North and South of Vietnam as well as within these regions. Finally, Section 6 offers some concluding remarks and policy recommendations.

#### 2. Paddy Production, Distribution and Marketing

Rice is the main staple in the Vietnamese diet. In the 1990s, household expenditure on rice was about 45 percent of total food expenditure in the rural areas and about 25 percent in the urban areas (Benjamin and Brandt, 2004). Rice policy issues ranging from measures to assist in the growing of paddy to the regulation of exports have considerable importance for agricultural researchers and policy makers. This section highlights some salient features of the Vietnamese paddy market.

#### 2.1 Natural Conditions for Paddy Production

About two thirds of Vietnamese farm households grow paddy to serve the high domestic demand for rice, exclusive of exporting requirement. Roughly 53 percent of Vietnam's agricultural land area, which is approximately 4 million hectares, is devoted to the growing of paddy (Minot and Golleti, 1999; FAOSTAT, 2006). Benjamin and Brandt (2004) summarize the state of Vietnamese paddy production as: "Rice production in Viet Nam is characterized by small irrigated farms, multiple cropping, labor-intensive practices, and growing use of inorganic fertilizer, though there are substantial regional differences". The planted area of paddy accounts for about 88 percent of the area planted of cereals in 2005 (Table 1). Although the area of paddy has been rising during 1993-2005, its share of total planted area of cereal has been declining in this period as a result of the higher growth rate of planted area of cereal compare with that of paddy.

The seasons for paddy growing vary across regions. In the North, there are generally two paddy crops annually: the Winter-Spring crop (planted in February and harvested in May/June) and the Summer-Autumn crop (July to October/November). In irrigated areas in the South, there are usually three paddy crops per year: the Summer-Autumn crop (April/May to August/September), the Autumn-Winter crop (August to November/December) and the Winter-Spring crop (November/December-February) (Minot and Golleti, 1999; Luu, 2003). In rain fed areas in the South, only a wet-season crop is produced, which is generally planted in July/August and harvested between October and February. By producing three crops staggered across the year, Vietnamese producers can supply paddy to the market more or less continuously throughout the year. Nationally, the Winter-Spring crop accounts for just below one-half of paddy production while the Summer-Autumn and Autumn-Winter crops each account for just over one-quarter (Figure 1 and Table 2).

Year	Agricultural land	Arable land	Planted area of cereal	Planted area of paddy	(4) as percentage of (3)
	(1)	(2)	(3)	(4)	(5)
1993	7087	5516	7058	6559	92.93
1994	7140	5464	7136	6599	92.47
1995	7079	5403	7324	6766	92.37
1996	7682	5554	7621	7004	91.91
1997	7844	5668	7768	7100	91.39
1998	8055	5763	8016	7363	91.85
1999	8413	6000	8349	7654	91.68
2000	8780	6200	8399	7666	91.28
2001	9483	6649	8225	7493	91.10
2002	9455	6600	8323	7504	90.17
2003	9622	6680	8367	7452	89.07
2004	n.a	n.a	8438	7445	88.24
2005 (prel.)	n.a	n.a	8371	7326	87.52

Table 1: Planted Land of Paddy (Thousand Hectares)

Source: GS0 (2006).



#### Figure 1: Paddy Production by Season in 1995-2005

Source: GSO (2006). Note: Figures for 2005 are preliminary

Year	Total		Of which (%)		
	(thousand tons)	Winter-Spring	Summer-Autumn	Autumn-Winter	
1993	22836.5	39.57	24.67	35.77	
1994	23528.2	44.66	24.14	31.20	
1995	24963.7	43.01	26.04	30.95	
1996	26396.7	46.25	26.06	27.69	
1997	27523.9	48.36	24.12	27.52	
1998	29145.5	46.52	25.81	27.67	
1999	31393.8	44.92	27.90	27.18	
2000	32529.5	47.87	26.51	25.62	
2001	32108.4	48.19	25.94	25.87	
2002	34447.2	48.54	26.67	24.79	
2003	34568.8	48.66	27.19	24.14	
2004	36148.9	47.24	28.86	23.90	
2005 (prel.)	35790.8	48.43	29.10	22.48	
Average		46.33	26.39	27.29	

#### Table 2: Paddy Production by Season, 1995-2005

Source: GSO (2006).

#### 2.2. The Vietnamese Paddy Policy

As the main Vietnamese staple food, rice has attracted significant attention from political leaders. Since the introduction of the contract system in 1980 (Directive 100), numerous reforms have been implemented, bringing about wider and deeper liberalization of Vietnam's paddy market, and a rise of around 12 million tons in paddy production from 1995 to 2005 (Figure 2). The main policy changes, which have affected the paddy market since 1980 include:

- The introduction of the *Doi moi* process in late 1986, which marked the beginning of a transition to a more market-oriented economy in Vietnam.
- The liberalization of market was gradually implemented from 1991, characterized by:
  - Exemption of duties on imported inputs for producing exported rice and the reduction of tariff on rice export from ten to one percent in 1991.
  - Extension of authority to import fertilizer from central to provincial SOEs with foreign exchange sources of revenues also in 1991. However, the imposition of quotas on, and licensing import of, fertilizer to selected enterprises remained until 2001.

- Loosening of restrictions and internal barriers to trade in rice between the northern and southern Vietnam was enacted in 1997 (Decree No. 140/TTg).
- A gradual lifting of the rice export quota, from under one million ton in 1992 to 4.5 million tons in 1998.
- Abolition of quotas on rice export and fertilizer imports, together with support provided for the rice producers and exporters, in 2001. Since then the free import of fertilizer has supplemented the insufficient amount produced by the domestic fertilizer industry, leading to a decrease in fertilizer price and a rise in the supply of fertilizer. This liberalization of fertilizer import and rice export consequently raised the paddy production and export volume.
- Land reform Resolution 5 (1993) extended land use right, creating incentives in paddy cultivation. The changes included (1) extension of tenure to 20 years for annual crop land and 50 years for perennials; (2) land user's right to exchange, transfer, lease, inherit and mortgage land.
- The declaration of the Water Law in 1999, which established a water allocation system using licenses and permits administered by Ministry of Agriculture and Rural Development (MARD) and the provincial governments. In 2002, the central management of water resource was assigned to the Ministry of Natural Resources and Environment, while the National Water Resources Council together with basin-level committees was set-up to manage and allocate water in the Red River Delta (RRD), Mekong River Delta (MRD) and Dong Nai basins. All these water legislations facilitated the intensification of crops from 1 to 2 or 3 crops per year.

In addition, a gradual improvement of market infrastructure and Vietnam's participation in international organizations has generally improved the trading environment.

#### 2.3. Paddy Production and Consumption

Vietnam's paddy production has been rising continuously during 1995-2005 with an average annual growth rate of 3.7 percent (Figure 3 and Table 3). In addition to the incentives from the rice market liberalization, this growth is attributed to an increase of 2.9 percent in yield of paddy (productivity) and a weak expansion of planted areas devoted to paddy, at the rate of about 0.8 percent per annum.



Figure 2: Trend of Paddy Production by Region in 1995-2005

Source: GSO (2006). Note: Figures for 2005 are preliminary.

Vietnam has two main granaries: the MRD in the South and the RRD in the North. These two deltas, which account for 42.5 percent of the population, supply rice to the remaining five regions (the North Uplands, North Central Coast, South Central Coast, Central Highlands, and Southeast). The MRD, the so-called "rice bowl of Vietnam", supply more than a half (52 percent) of national paddy output, followed by the RRD that produces around 20 percent of the total. None of the other five regions accounts for more than 9 percent of national paddy production (Table 3).

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Average
As proportion of total paddy production (%)												
Vietnam	100	100	100	100	100	100	100	100	100	100	100	
Red River Delta	20.4	20.3	20.7	20.7	20.3	20.2	20.0	19.6	18.8	18.6	17.3	19.7
North East	5.8	6.1	6.3	6.1	6.1	6.3	7.0	6.9	7.2	6.9	7.1	6.5
North West	1.3	1.2	1.3	1.1	1.2	1.2	1.4	1.3	1.4	1.5	1.5	1.3
North Central Coast	8.6	7.8	9.1	8.0	8.4	8.7	9.2	9.2	9.3	9.3	8.8	8.8
South Central Coast	5.7	6.0	5.8	5.4	5.4	5.2	5.3	5.0	5.4	5.2	5.0	5.4
Central Highlands	1.7	1.6	1.8	1.5	1.6	1.8	2.0	1.8	2.2	2.2	2.0	1.8
South East	5.1	4.3	4.3	4.2	5.0	5.2	5.2	4.9	5.0	4.9	4.5	4.8
Mekong River Delta	51.4	52.8	50.8	53.0	51.9	51.3	49.8	51.4	50.7	51.4	53.7	51.7
Growth rate of paddy proc	luction (%	6)										
Vietnam		4.9	4.2	6.0	8.6	3.6	-1.3	7.3	0.4	4.6	-1.0	3.7
Red River Delta		4.6	5.9	6.1	6.8	3.2	-2.5	5.2	-3.9	3.4	-7.6	2.0
North East		8.9	7.8	3.2	8.2	8.0	9.0	5.5	4.2	0.6	1.9	5.7
North West		-4.0	9.4	-3.8	12.0	8.4	9.2	3.8	6.7	12.4	-0.5	5.2
North Central Coast		-4.9	22.6	-7.2	13.7	7.2	5.1	6.4	2.1	4.9	-6.3	4.0
South Central Coast		10.7	0.8	-1.0	8.9	-1.3	1.5	0.2	9.8	0.7	-6.1	2.3
Central Highlands		-0.5	13.6	-10.1	17.4	14.5	10.1	-6.1	23.3	4.5	-8.6	5.2
South East		-12.2	6.0	1.9	31.2	6.2	0.1	-0.1	3.8	2.3	-9.2	2.5
Mekong River Delta		7.7	0.2	10.6	6.4	2.5	-4.2	10.7	-1.0	5.9	3.6	4.1

Table 3: Paddy Production by Region, 1995-2005

Notes: Figures for 2005 are preliminary.

Source: Computed from GSO (2006).

	1993-1995	1996-1998	1999-2001	2002-2003	Average
Paddy (thousand tons)	69127.8	82557.4	89073.8	60064.9	
Production	71328.6	83066.1	96031.7	69016.0	
Imports	17.3	2.0	11.8	64.0	
Stock Change	-2218.1	-510.7	-6969.7	-9015.2	
Provided for:					
Exports	8372.1	15484.1	17749.6	10687.8	
Domestic Supply	60755.7	67073.3	71324.3	49377.1	
In which:					
Feed	1132.6	1270.9	1366.8	1168.2	
Seed	2139.9	2492.0	2881.0	2070.5	
Waste	4865.0	5670.0	6549.6	4714.4	
Food Manufacture	335.9	459.0	608.6	150.7	
Food	51957.9	56802.2	59481.2	40957.4	
Other Uses	324.6	379.2	437.0	315.9	
Paddy (%)	100.0	100.0	100.0	100.0	100.0
Production	103.2	100.6	107.8	114.9	106.6
Imports	0.0	0.0	0.0	0.0	0.0
Stock Change	-3.2	-0.6	-7.8	-15.0	-6.7
Provided for:					
Exports	12.1	18.8	19.9	17.8	17.2
Domestic Supply	87.9	81.2	80.1	82.2	82.9
In which:					
Feed	1.9	1.9	1.9	2.4	2.0
Seed	3.5	3.7	4.0	4.2	3.9
Waste	8.0	8.5	9.2	9.6	8.8
Food Manufacture	0.6	0.7	0.9	0.3	0.6
Food	85.5	84.7	83.4	83.0	84.1
Other Uses	0.5	0.6	0.6	0.6	0.6

Table 4: Distribution of Paddy Production, 1993-2003

Notes: Data for 2004 and 2005 are not available.

Source: Computed from FAOSTAT Data, 2006.

In the RRD, the average annual growth rate of paddy production was 2 percent from 1995 to 2005. That is 2.1 percentage points lower than the growth rate for paddy in the MRD. However, the MRD faces higher variability in production than the RRD (Figure 3). Among the provinces in the RRD, Thai Binh and Nam Dinh rank first and second in paddy production, each accounting for around 3 percent of national output and 15 or 16 percent of RRD's production (Figure 4). They are followed by Ha Tay and Hai Duong with about 14 and 13 percent of RRD's paddy production, respectively. The other seven provinces in the RRD produce less than 8 percent of national paddy production.



Figure 3: Growth Rate of Paddy Production by Region, 1995-2005

Source: GSO (2006) Note: Figures for 2005 are preliminary



Figure 4: Paddy Production in Red River and Mekong River Deltas, 1995-2005

*Note:* Figures for 2005 are preliminary

The leading paddy growing provinces in the MRD include An Giang, Dong Thap, Kien Giang and Can Tho whose shares of national output are each 6 to 7 percent (accounting for 11 to 14 percent of the MRD's annual paddy production). They are followed by Long An, Soc Trang and Tien Giang, which together account for almost 5 percent of national production. This is equivalent to more than 9 percent of the MRD's annual production (Figure 4).

Besides the domestic demand, Vietnam also supplies rice to the international market. About 83 percent of total annual paddy production is consumed domestically while the other 17 percent is

exported. As can be seen from Table 4, the main portion of domestic consumption (about 84 percent) is used for food, with the remaining 16 percent accounted for by waste, seed and feed. Since the mid 1990s, Vietnam has become a major rice exporting country, with the volume of rice exports rising from 2.5 million tons in 1993 to 5.8 million tons in 2003 (Figure 5). In terms of rice export volumes, Vietnam is now the second largest rice exporter in the world after Thailand (Figure 6). However, in terms of value, Vietnam's rice exports rank fourth or fifth in the world market due to the low quality of its rice exports and the limited value-added to paddy (Figure 6).



#### Figure 5: Vietnam's Export of Rice in Paddy Equivalent







Source: FAOSTAT data, 2006

#### 2.4. Paddy Marketing

Before sale to consumers as milled rice, paddy goes through a process that involves many stages, including checking and sorting, drying and storing, milling, polishing and packaging, storage and transportation (Luu, 2003). There are various channels used for marketing paddy and rice in Vietnam (Figure 7). According to Minot and Goletti (1999), more than two thirds of paddy is sold to the assemblers, of whom 95 percent are private firms. Most of the remainder goes to small-scale millers for farmers' home consumption, while a small share is bought directly by larger millers. The assemblers then either have the paddy custom milled or sell paddy to medium and large-scale millers. The assemblers and millers then sell the milled rice on to the wholesalers and retailers. State-owned enterprises (SOEs) play a relatively minor role in the within province marketing of paddy and rice but control the long-distance trade in rice ad export market. SOEs consist of provincial and regional food companies (of which Vinafood I in the North, and Vinafood II in the South are the most important). Until recently only the SOEs were permitted to export rice to other countries, and they still dominate the export trade.



Figure 7: Vietnam's Paddy and Rice Marketing Channels

Source: Drawn by the authors, based on Luu (2003) and Minot and Golletti (1999)

#### 3. Modelling and Testing for Market Integration with Transfer Costs

Tests of spatial market integration using time series data on food price dates back to the 1960s. The first tests arose from Lele's (1967) and Jones' (1968) analyses of staple food prices in India and Nigeria, respectively. The two authors estimated the contemporaneous correlations between price series in two markets in different locations. When the correlations were greater than 0.7 or 0.8, they concluded that price movements were close enough for the two markets to be considered as being spatially integrated. The cut-off point used for deciding if markets were spatially integrated was, however, arbitrary and no consideration was given to whether or not the price series were stationary or non-stationary. Nonetheless, many of the subsequent (more rigorous) econometric models used for testing spatial market integration rely on the same idea of testing how closely prices move together. In particular, models of the Law of One Price (Isard, 1977; Richardson, 1978) and the Ravallion model (Ravallion, 1986) are both extensions of the correlation idea. Cointegration tests have also been used to test for the co-movement of food prices and long-run market integration (Alexander and Wyeth, 1991; Dawson and Dey, 2002).

A growing body of the spatial market integration literature stresses the importance of transfer costs. Basically, transfer costs introduce a wedge between prices at separate locations, resulting in a non-linear relationship between such pairs of prices. The parity bounds model developed by Sexton, Kling and Carman (1991) and Baulch (1997) explicitly take account of the non-linear price relationship in spatially distributed markets that is caused by transfer costs. Furthermore, much recent research focuses on an explicit modeling of threshold effects when testing the Law of One Price. See, for example, Goodwin and Piggott (2001), Meyer (2004), and Sarno, Taylor and Chowdhury (2004).

Transfer costs are important both for modeling spatial market integration as they imply a potential bias in estimators based on linear models. Consider, for example, a relationship between two paddy markets, 2 and 1. In the absence of transfer costs, spatial market integration will imply that the paddy prices in the two markets should be equal, so  $p_1 = p_2$ . In Figure 8, such a market integration situation is illustrated by the 45-degree line. Now, if there are transfer costs of tc per kilogram of paddy, there need not be any trade between the markets when  $p_2 - tc < p_1 < p_2 + tc$  because the transfer costs make inter-market trade unprofitable. This "no-trade" region suggests that prices respond differently in the three regimes defined by the upper and lower thresholds,  $p_2 + tc$ ,  $p_2 - tc$ . If  $p_1$  is above  $p_2 + tc$  (for example, point A in Figure 8) one would expect a trade flow from market 2 to market 1, leading to a price increase in market 2 and a price decrease in market 1. Likewise, if  $p_1$  is less that  $p_2 - tc$  (for example, point B in Figure 8) trade between the markets should lead to a price decrease in market 2 and a price increase in market 1. These effects are not changed by the presence of transfer costs. However, prices will only adjust when trade is profitable. When  $p_1$  is between the  $p_2 - tc$  and  $p_2 + tc$ lines in Figure 8 (for example, point C), then there is no trade inducing market forces to equate the paddy prices in the two markets. We describe this as the no-trade regime. It follows that the overall correlation between the prices may be low if prices are often in this no-trade regime. Hence, if we use a linear model when there are positive transfer costs, we are likely to underestimate the price correlations in the regimes in which trade is profitable because these correlations are confounded by the possible lack of correlation in the no-trade regime.



#### Figure 8: Prices in Two Markets with Transfer Costs

The main problem with threshold models of spatial market integration is that in the absence of information on transfer cost the estimation procedure is rather complicated, particularly with two thresholds.<sup>1</sup> To overcome this problem, in this chapter we use outside information on transfer costs to formulate a threshold variation of Ravallion's well-known dynamic model of spatial market integration (Ravallion, 1986). Specifically, we use fixed and known threshold values computed from additional information (collected from traders and transporters) about transfer costs to model the three regimes described above. The remainder of this section shows how this model is formulated and presents the testing sequence we use to analyze the various forms of spatial market integration.

Consider Ravallion's model of market integration in which price in market 1 ( $p_{1t}$ ) is conditional on the price in market 2 ( $p_{2t}$ ), lagged prices ( $p_{1t-i}$  and  $p_{2t-i}$ ) and a set of seasonal dummies ( $S_t$ ):<sup>2</sup>

$$p_{1t} = d + a_1 p_{1t-1} + a_2 p_{1t-2} + b_0 p_{2t} + b_1 p_{2t-1} + b_2 p_{2t-2} + \phi' S_t + \mathcal{E}_t$$
(1)

We generalize this model by allowing the constant, the autoregressive parameters and the distributed lag parameters to vary across regimes in the following way:

$$p_{1t} = d^{(s)} + a_1^{(s)} p_{1t-1} + a_2^{(s)} p_{1t-2} + b_0^{(s)} p_{2t} + b_1^{(s)} p_{2t-1} + b_2^{(s)} p_{2t-2} + \phi' S_t + \varepsilon_t$$
<sup>(2)</sup>

in which  $\theta^{(s)} = (d^{(s)}, a_1^{(s)}, a_2^{(s)}, b_0^{(s)}, b_1^{(s)}, b_2^{(s)})$  are regime dependent parameters defined by

$$\theta^{(s)} = \begin{cases} \theta^{(1)} & \text{if} (p_{1t-l} - p_{2t-l}) \ge tc_1 \\ \theta^{(0)} & \text{if} tc_2 < (p_{1t-l} - p_{2t-l}) < tc_1 \\ \theta^{(2)} & \text{if} tc_2 \ge (p_{1t-l} - p_{2t-l}) \end{cases}$$
(3)

where the relationship between the inter-market price spread at lag *l* and the transfer costs ( $tc_1$ ,  $tc_2$ ) determines the regime. For simplicity, we assume that the additive seasonal component in the price series is the same in all three regimes, and we also assume the innovations are  $iid(0, \sigma^2)$  across the regimes.

<sup>&</sup>lt;sup>1</sup> See Hansen (1996, 1997) for a general estimation procedure of threshold models and Hansen and Seo (2002) for estimation of error correction models with threshold effects.

<sup>&</sup>lt;sup>2</sup> An autoregressive, distributed lag model with two lags appears to give a good description of the paddy price series in the subsequent analysis. Therefore, we present the model with two lags rather than a more general model having, say, k lags.

The model in (2) can be reformulated as an error correction model allowing for non-stationary and cointegrated prices in the two markets

$$\Delta p_{1t} = \gamma^{(s)} \Delta p_{2t} + \delta_1^{(s)} \Delta p_{1t-1} + \delta_2^{(s)} \Delta p_{2t-1} + \alpha^{(s)} (p_{1t-1} - \beta^{(s)} p_{2t-1} - \mu^{(s)}) + \phi' S_t + \varepsilon_t$$
(4)

and the parameters in (2) and (4) are related by:  $\mathbf{r}^{(s)} = \mathbf{L}^{(s)} \mathbf{S}^{(s)} = \mathbf{r}^{(s)} \mathbf{S}^{(s)} = \mathbf{L}^{(s)}$ 

$$\gamma^{(s)} = b_0^{(s)}, \, \delta_1^{(s)} = -a_2^{(s)}, \, \delta_2^{(s)} = -b_2^{(s)}$$

$$\alpha^{(s)} = -(1 - a_1^{(s)} - a_2^{(s)}), \, \beta^{(s)} = \frac{b_0^{(s)} + b_1^{(s)} + b_2^{(s)}}{1 - a_1^{(s)} - a_2^{(s)}}, \, \mu^{(s)} = \frac{d^{(s)}}{\alpha^{(s)}}$$
(5)

It is worth noting that the error correction term in (4) is a "generalized" inter-market price spread  $w_t = p_{1t} - \beta^{(s)} p_{2t} - \mu^{(s)}$ , which simplifies to the standard inter-market price spread when  $\beta^{(s)} = 1$ .

The threshold error correction in equation (4) can be used to test most of the interesting hypotheses about spatial market integration including the Law of One Price and Ravallion's tests for market segmentation and long-run market integration. To ensure valid inference, our first step is to test for stationarity of the individual price series and of the generalized inter-market price spread.<sup>3</sup> Conditional on the existence of cointegration, we then test the hypotheses of interest for spatial market integration using the threshold error correction model.

The first hypothesis to be tested is one of *market segmentation*. Based on Ravallion (1986), we define market segmentation as a model in which the price in market 2 does not Granger-cause the price in market 1 and, in addition, that there is no instantaneous correlation between the price series—in any regime. Hence, in terms of parameter restrictions we say that the markets are segmented if in equation (4):

$$H_0^{(1)}: \alpha^{(s)} = 0, \, \delta_2^{(s)} = 0, \, \gamma^{(s)} = 0, \quad \text{for all } s = 0, 1, 2.$$
(6)

If market segmentation is rejected, we next test for the number of thresholds. This is done in three steps. Specifically, we formulate and test the three hypotheses

$$H_0^{(2a)}: \theta^{(1)} = \theta^{(0)}, \quad H_0^{(2b)}: \theta^{(2)} = \theta^{(0)}, \quad H_0^{(2c)}: \theta^{(1)} = \theta^{(2)} = \theta^{(0)}.$$
(7)

If the first two hypotheses are rejected we infer that there are two thresholds. In contrast, if all three hypotheses are accepted we infer that the model is linear.

Once the number of thresholds is established, we can further reformulate the error correction model to allow for freely varying long-run parameters in each regime

$$\Delta p_{1t} = \gamma^{(s)} \Delta p_{2t} + \delta_1^{(s)} \Delta p_{1t-1} + \delta_2^{(s)} \Delta p_{2t-1} + \alpha^{(s)} p_{1t-1} + \pi^{(s)} p_{2t-1} + d^{(s)} + \phi' S_t + \varepsilon_t$$
(8)

Subsequently, we test for long-run market integration, which is the hypothesis

$$H_0^{(3)}: \alpha^{(s)} = -\pi^{(s)}$$
 (equivalent to  $\beta^{(s)} = 1$ ) (9)

in equation (8). This is equivalent to testing  $\beta^{(s)} = 1$  in equation (4). Clearly, the number of hypotheses tested depends on the number of thresholds, so our tests for long-run market integration are regime dependent. Note that with two thresholds, the information about the long run parameters may be weak in particular in the inner (no-trade) regime.<sup>4</sup>

<sup>&</sup>lt;sup>3</sup> We test for stationarity of the price series and the price spread in the simplest possible way using the ERS unit-root test (Elliot, Rothemberg and Stock, 1996). As we do not include thresholds when testing stationarity of the generalized inter-market price spread, we are effectively assuming the long run parameters are constant across regimes when testing for stationarity. While the assumption of regime independent long run parameters is a restriction on the model, we consider this to be a quite plausible initial restriction and later allow for differences in the long-run parameters.

<sup>&</sup>lt;sup>4</sup> It should also be noted that the term 'long-run market integration' may be somewhat misleading in this model as acceptance of the hypothesis in some regimes does not imply that prices will eventually be equal in the absence of shocks. However, if the hypothesis is accepted in all regimes, then the standard notion of long-run market integration applies.

If long run integration is accepted in a regime, we impose the restriction by using the lagged intermarket price spread as the error correction term (*i.e.*,  $\alpha^{(s)}p_{1t-1} + \pi^{(s)}p_{2t-1}$  is replaced by  $\alpha^{(s)}(p_{1t-1} - p_{2t-1})$  in equation (8). If long-run market integration is not accepted it makes little sense to carry on testing stronger versions of the Law of One Price. Hence, the following hypotheses and tests are only reported if the hypothesis of long-run market integration is accepted in at least one regime.

Before testing for the strong version of the Law of One Price we look at a weaker restriction, which is akin to the weak market efficiency hypothesis in the financial markets literature. This is done by formulating a hypothesis of *common dynamics*. We define common dynamics to be the restriction

$$H_0^{(4)}: \delta_1^{(s)} = \delta_2^{(s)} \tag{10}$$

If the hypothesis is accepted, the only past information needed to predict the price in market 1 is the inter-market price spread in previous periods. Adding information on past values of individual prices will not improve the prediction of the price in market 1.

Finally, given acceptance of long-run market integration and common dynamics it is interesting to test for the strong version of the *Law of One Price*. The hypothesis is formulated as

$$H_0^{(5)}: \gamma^{(s)} = 1, \text{ and } \delta_1^{(s)} = \delta_2^{(s)}$$
 (11)

This involves testing jointly that the prices in the two markets have common dynamics and that the conditional correlation between the two prices is one. In this case, a unit change in prices in market 2 will be associated with a one unit change in prices in market 1 within the current time period. This is similar to Ravallion's (1986) test for short-run market integration (strong form) and implies that the Law of One Price holds contemporaneously.<sup>5</sup>

The test for Law of One Price is the final test in the sequence. If this test is not rejected in the regimes with trade options, we consider the markets to be fully integrated even if the hypothesis is rejected in the no-trade regime.

$$\Delta p_{1t} = \Delta p_{2t} + \delta_1^{(s)} \Delta (p_{1t-1} - p_{2t-1}) + \alpha^{(s)} (p_{1t-1} - p_{2t-1} - \mu^{(s)}) + \phi' S_t + \varepsilon_t$$

<sup>&</sup>lt;sup>5</sup> If the Law of One Price is not rejected in a regime, the model for that regime reduces to

and this implies that the loading parameter  $\alpha^{(s)}$  is no longer of "special interest" for describing the dynamics of price adjustment. The differences between prices in markets 1 and 2 will, in this regime, be a white noise process, given by  $\mathcal{E}_{r}$ .

#### 4. Data on Paddy Prices, Transfer Costs and Time Series Properties

#### 4.1. Data

The paddy price data used in this chapter were provided by Information Center for Agriculture and Rural Development (ICARD) and can be publicly accessed online.<sup>6</sup> Prices of paddy (and some other agricultural commodities) are collected on Fridays at public markets in sixteen provinces. Based on the weekly data, monthly price series are computed by taking averages. It is worth considering the appropriateness of using monthly series in testing spatial market integration. Luu (2003) argues that average monthly data is inappropriate when analyzing rice market integration in the Mekong River Delta because these prices do not reflect the daily prices on which traders make their arbitrage decisions. However, in the weekly price series we have looked at there are several long periods in which prices do not change at all in almost every market. Such constancy of prices will invalidate a statistical analysis that is based on an assumption of independent and identically distributed innovations that follow a continuous distribution. Moreover, the use of weekly data is problematic due to the need to interpolate numerous missing values. For these reasons we have chosen to analyze the monthly data instead of the weekly data. However, since one month appears too long a time lag for traders to make their arbitrage decisions we construct the regime dummies using a lag of half a month.<sup>7</sup>

For paddy prices, monthly data are fully available across eight markets from January, 1993 to May, 2006. Given the fact that inflation has changed substantially over the period, we adjust for its effects by deflating nominal paddy prices using the Consumer Price Index. Accordingly, all paddy prices to be used in the analysis were converted into constant January 1993 prices. Four out of the eight markets are located in the Red River Delta (Ha Noi, Bac Ninh, Nam Dinh, and Thai Binh). The other four markets are located in the Mekong River Delta (Can Tho, An Giang, Tien Giang, and Ca Mau). As can be seen in Figure 9, all these provinces cluster round the 1A National Road, which is the main road connecting the North, Center and South of Vietnam. National roads also connect Ha Noi with Bac Ninh, Nam Dinh and Thai Binh and Can Tho with An Giang and Tien Giang. In both the Red River and Mekong River deltas, rivers are also used for the transportation of paddy: one of the Mekong River's branches, the Hau Giang, connects An Giang with Can Tho.

<sup>&</sup>lt;sup>6</sup> ICARD is an institution under the Ministry for Agriculture and Rural Development (MARD). See the website: http://www.agroviet.gov.vn.

<sup>&</sup>lt;sup>7</sup> We do this by taking an average of the inter-market price spreads at time *t* and *t-1. i.e.*,  $\lambda(p_{1t} - p_{2t}) + (1 - \lambda)(p_{1t-1} - p_{2t-1})$  where  $\lambda = 0.5$ . Setting  $\lambda$  equal to 0.25 or 0.75 produces similar results.



Figure 9: Location and Transportation Links between the Eight Paddy Markets

*Source:* Drawn by the authors using ArcGIS

In addition to price series, we also need estimates of transfer costs to determine the thresholds used in the estimation of the threshold error correction model defined by equations (4) and (8). Transfer costs represent the total observed costs of moving paddy from one market to another, and include both fixed costs and variable costs elements. Data on these fixed and variable costs were collected through interviews with traders and lorry drivers. The fixed cost element comprises loading and unloading costs plus specific trade taxes (where applicable). The variable cost element comprises pure freight costs, which vary with the distance traveled and road conditions, and *ad valorem* trade taxes. From our interviews with traders, we have estimated the unit cost of transporting one kilogram of paddy over one kilometer. We assume that these unit costs are constant in real terms and do not vary according to the distance between markets.<sup>8</sup> So to estimate the transportation costs between two markets, we simply multiply the unit costs of transportation by the distance between markets. Finally, since it is

<sup>&</sup>lt;sup>8</sup> In theory, unit transportation costs should decline with distance traveled. However, our discussions with traders and transporters indicate that this is a very imprecise relationship with freight rates often remaining constant between nearby markets.

rice rather than paddy which is usually transported between provincial markets, we also have to adjust transfer costs by the milling ratio between paddy and rice. Accordingly, our estimates of transfer costs between markets *i* and *j* ( $tc_{ij}$ ) are based on the following expression:

$$tc_{ij} = \frac{20 + 0.45 \times D_{ij}}{0.65} \tag{12}$$

where 20 (Dong) is the fixed cost of loading and unloading one kilogram of rice, 0.45 (Dong) is the unit cost of transporting 1 kilogram of paddy for 1 kilometer,  $D_{ij}$  is the distance following the main road between market *i* and market *j* and 0.65 is the milling (rice-paddy conversion) ratio.

Table 5 presents our estimated transfer costs between the market pairs in our analysis. Within each region (North or South) we assume that transfer costs from market *i* to market *j* is the same as from market *j* to market *i*. However, our discussions with traders suggest that this assumption does not hold between markets in the North and South of Vietnam, since the demand for moving goods from the South to the North is substantially higher than in the opposite direction. Our discussions with traders indicate that premium on moving produce from the South to the North is approximately 25 percent, so this is added to our estimates of transfer costs between Can Tho and Ha Noi. To be consistent with the paddy price series, we have then deflated our transfer cost estimates to constant January 1993 terms using the Consumer Price Index.

Market 1	Market 2	Distance (km)	Estimated Transfer Cost
Ha Noi	Bac Ninh	25	23
Ha Noi	Nam Dinh	85	43
Ha Noi	Thai Binh	95	46
Ha Noi	Can Tho	1794	606
Can Tho	Ha Noi	1794	758
Can Tho	An Giang	65	36
Can Tho	Tien Giang	115	53
Can Tho	Ca Mau	213	85

Table 5: Estimated Transfer Costs Between Market Pairs

Notes: The transfer costs are measured in Dong per kg in January 1993 prices.

#### 4.2. Descriptive Statistics and Tests for Stationarity and Cointegration

Table 6 shows means, medians, and other basic statistics of the monthly price series. We note that, on average, prices in the northern provinces are lower than those in the southern provinces. Among the eight markets in our sample, An Giang and Can Tho have the lowest paddy prices. This low level of prices may well reflect that An Giang and Can Tho are the largest paddy producers in Vietnam. Another interesting result in Table 6 is that, on average, paddy prices in Ha Noi were lower than in Bac Ninh. To this end it should be noted that the median price in Bac Ninh is lower than the median price in Ha Noi. Figures 10 and 11 show the monthly paddy prices in the four provinces in the North and the four provinces in the South, respectively. Based on the plots, the overall impression is that there is a fairly high co-movement in prices in the two regions. Furthermore, there is no clear trend in the series; on the other hand there is also no clear mean reversion tendency.
Market	Market Mean		Std. dev.	Min	Max
		North	ו		
Ha Noi	1348.89	1363.12	219.95	836.11	1934.56
Bac Ninh	1375.96	1350.53	232.18	1021.99	2063.53
Nam Dinh	1308.69	1302.00	222.92	935.51	1930.30
Thai Binh	1286.03	1267.31	212.27	919.78	1862.91
		South	'n		
Can Tho	1090.79	1093.13	137.32	765.33	1412.56
An Giang	1083.93	1083.61	155.86	716.89	1487.86
Tien Giang	1112.76	1112.13	149.78	755.10	1560.78
Ca Mau	1101.36	1106.82	166.43	729.56	1668.21

Table 6: Descriptive Statistics for the Paddy Prices

Notes: There are 161 observations in the sample (1993m1-2006m5). Prices are measured in terms of Dong per kg in January 1993 prices.



Figure 10: Monthly Paddy Prices in the Northern Provinces (1993m1-2006m5)

Source: Drawn by the authors using price data from the Information Center for Agriculture and Rural Development



Figure 11: Monthly Paddy Prices in the Southern Provinces (1993m1-2006m5)

Source: Drawn by the authors using price data from the Information Center for Agriculture and Rural Development

We next examine the time series properties of the prices by testing for stationarity using Elliott, Rothenberg, and Stock's (1996) modified Dickey-Fuller test for unit roots.<sup>9</sup> The ERS test involves expressing a time series  $y_t = y_1, ..., y_T$  as:

$$y_t = \tau_t + \eta_t,$$

$$\eta_t = \rho \eta_{t-1} + \nu_t$$
(13)

where  $\tau_t$  is a deterministic component,  $v_t$  is a stationary process with zero mean, and  $\rho$  is the autoregressive parameter. The null hypothesis that  $\rho = 1$  implies that the series  $y_t$  has a unit root (i.e., is integrated of order one, I(1)) and it is tested against the alternative  $|\rho| < 1$ , implying that  $y_t$  is stationary.

The ERS procedure works as follows. First, a local unit root is assumed by fixing the parameter  $\overline{\rho} = (1 + \overline{c} / T)$ .<sup>10</sup> The parameter is used to form quasi-differences of the series:  $\{y_1, y_2 - \overline{\rho} y_1, y_3 - \overline{\rho} y_2, ..., y_T - \overline{\rho} y_{T-1}\}$  and the quasi-differences are regressed on corresponding quasi-differences of the deterministic components (constant, trend, and so on). Subsequently, a standard ADF-test is used to test if the non-deterministic component of the series ( $\eta_i$ ) has a unit root.

Table 7 reports the results of testing for unit roots in each price series using the ERS test with a

<sup>&</sup>lt;sup>9</sup> This test is also known as the DFGLS test. Basically, this test follows the augmented Dickey-Fuller procedure, but, as stated in Baum and Sperling (2001), it has better power than the ADF-test in small samples.

<sup>&</sup>lt;sup>10</sup> In most cases,  $\pi$  is zero, a constant or a linear trend.  $\overline{c}$  is set to -13.5 if the model contains a linear trend, and -7 if it does not.

trend as the dominating deterministic component. The number of lags is chosen using the Schwarz information criterion. All price series in the northern markets have unit roots in levels but not in first differences, hence they are I(1). However, in the southern markets the price series from An Giang, and Tien Giang do not seem to have unit roots when tested at the 5 percent level of significance though they do have unit roots at the 1% level of significance. The test statistic for Ca Mau lies on the boundary of accepting or rejecting the unit root at the 5 percent level. Like the North, null hypothesis of a unit root is rejected for the first differences of the prices in the South. In sum, we conclude that, all the price series are integrated of order 1.

Next we test for stationarity of residuals from pair-wise Engle-Granger-type cointegrating regressions and stationarity of the inter-market price spreads (Table 8). We expect to reject the null hypothesis of a unit root in the residuals for the assumptions underlying our general threshold error correction model to hold. In general, we can strongly reject the hypothesis of a unit root and conclude that there is cointegration between all market pairs. However, we cannot reject the null hypothesis of non-stationarity for the residuals of regressing Can Tho on Tien Giang. For the inter-market price spreads, which are used for the error correction terms in the restricted version of our model, we can also reject a unit root for all market pairs at the 5 percent level.

	First diff	erences	Levels		
Market	Optimal lag	Test statistic	Optimal lag	Test statistic	
North	· · ·				
Ha Noi	1	-10.31	2	-2.21	
Bac Ninh	1	-8.42	1	-2.34	
Nam Dinh	1	-8.91	1	-2.35	
Thai Binh	1	-8.97	1	-2.34	
South					
Can Tho	4	-3.06	2	-2.48	
An Giang	1	-10.13	1	-3.22	
Tien Giang	1	-9.75	1	-2.96	
Ca Mau	3	-3.32	1	-3.52	

### Table 7: Unit-Root Tests

*Notes:* The test for unit-roots is the Elliot, Rothenberg and Stock (1996) test with seasonal dummies included. The 1%, 5%, and 10% critical values for levels are -3.51, -2.91, and -2.63, respectively. The 1%, 5%, and 10% critical values for first differences are -2.59, -2.04, and -1.72, respectively.

Residuals from										
Cointegration regressions Inter-market price spread										
Market pair	Optimal lag	Test statistics	Optimal lag	Test statistics						
Ha Noi - Can Tho	1	-3.37	1	3.41						
North										
Ha Noi - Bac Ninh	2	-3.69	2	-3.84						
Ha Noi - Nam Dinh	1	-5.26	1	-5.33						
Ha Noi - Thai Binh	1	-4.19	1	-4.45						
South										
Can Tho - An Giang	2	-5.78	2	-5.65						
Can Tho - Tien Giang	3	-2.42	3	-2.38						
Can Tho - Ca Mau	1	-4.97	1	-5.56						

### **Table 8: Pair-wise Cointegration Tests**

*Notes:* The test for unit root is the Elliot, Rothenberg and Stock (1996) test with seasonal dummies included. The 1%, 5%, and 10% critical values for the residuals from the cointegration regressions are -3.43, -2.86, and -2.57, respectively. The 1%, 5%, and 10% critical values for the inter-market price spread are -2.59, -2.04, and -1.72, respectively.

### 5. Tests for Market Integration

In this Section, we apply the sequential approach to testing for market integration discussed in Section 3 to the eight non-stationary and cointegrated paddy price series described above in order to investigate: (i) whether paddy markets in the North and the South of Vietnam are spatially integrated, (ii) if paddy markets within the North and within the South are integrated, and (iii) if price adjustment within regions are stronger (and faster) than between them.

To investigate whether paddy markets in the North and the South of Vietnam are spatially integrated, we choose Ha Noi to represent the North of Vietnam and Can Tho to represent the South. Ha Noi, the capital city of Vietnam, is the second largest city in Vietnam and it is located close to the centre of the Red River Delta. Can Tho is the major milling and exporting centre in the Mekong River Delta.<sup>11</sup> Having tested if Ha Noi and Can Tho are spatially integrated, we then move on to examining whether the other three markets in the North (Bac Ninh, Nam Dinh and Thai Binh) are spatially integrated with Ha Noi. Then we perform similar tests with Can Tho and the other three markets in the South (An Giang, Tien Giang and Ca Mau). Recall that only when market segmentation is rejected and long-run market integration is accepted, do we move on to testing for common dynamics and the Law of One Price.

Table 9 summarizes the test results for all market pairs. Two thresholds are indicated for all market pairs except for Ha Noi and Can Tho (where no thresholds are detected) and between Ha Noi and Bac Ninh (where there is no statistical difference between the parameters in the above and inside regimes). Market segmentation is also strongly rejected for all market pairs, except for Can Tho and Ca Mau, implying that paddy prices in the demand centers are influenced by paddy prices in the markets in the producing areas.

<sup>&</sup>lt;sup>11</sup> Ho Chi Minh City, the major commercial centre and largest city in Vietnam, would have been an alternative choice for the major centre of demand in the South. However, it is not located in the Mekong River Delta and the paddy price data available for Ho Chi Minh City are very incomplete.

Market pairs	# of thres- holds	Regime Frequencies	Long-run Integration	Common Dynamics	Law of One Price	γ	α
Ha Noi - Can Tho	0		A	R	R	0.410	-0.149
		[1.00]	[0.287]	[0.009]	[0.000]	(0.105)	(0.040)
North						, , ,	, ,
Ha Nai - Raa Ninh	4	Above	٨	٨	D	0.644	1 101
Tha NUI - Dac Millin	1	Inside	~	~	n	0.044	-1.101
		[0.50]	[0.592]	[0.095]	[0.000]	(0.073)	(0.356)
		Below	A	А	R	0.637	-0.606
		[0.50]	[0.071]	[0.632]	[0.000]	(0.082)	(0.096)
Ha Noi - Nam Dinh	2	Above	A	A	R	0.600	-0.437
		[0.42]	[0.853]	[0.446]	[0.000]	(0.084)	(0.105)
		Inside	A	A	A	1	
		[0.37]	[0.208]	[0.990]	[0.741]	()	
		Below	A	A	Α	1	
		[0.21]	[0.620]	[0.065]	[0.129]	()	
Ha Noi - Thai Binh	2	Above	A	A	R	0.495	-0.407
		[0.54]	[0.941]	[0.964]	[0.000]	(0.073)	(0.088)
		Inside	A	A	R	0.779	-1.394
		[0.32]	[0.438]	[0.871]	[0.012]	(0.074)	(0.151)
		Below	A	A	A	1	
<b>0</b> //		[0.14]	[0.629]	[0.512]	[0.203]	()	
South	•				-	0 700	4 000
Can Tho - An Giang	2	Above	A	A	R	0.733	-1.083
		[0.27]	[0.838]	[0.419]	[0.001]	(0.073)	(0.1/5)
			A	A [0.070]	K [0.040]	0.859	-1.357
		[0.57] Deleve	[0.125]	[0.870]	[0.049]	(0.057)	(0.117)
		Below	A [0.150]	K [0.000]	H [0.000]	0.868	-1.201
Can The Tion Ciona	0	[U. 10] Above	[0.156]	[0.039]	[0.039]	(0.010)	(0.158)
Can Tho - Tien Giang	2		A [0 907]	A [0 490]			
		[U.U0] Incido	[0.007]	[0.432] A	[0.009] D	()	0 5 9 7
			A [0.067]	A [0.064]		0.790	-0.307
		[U.UZ] Rolow	[U.UU7] A	[U.U04] ^	[0.000] D	(0.002) 0.000	(0.095)
			A [0 040]	A [0 770]		0.000	-0.947
Can Tho - Ca Mau	0		[0.240] Δ	Δ	[0.010] B	(0.000) 0 152	-0.864
Can Tho - Ca Mau	2	190 01	n 10 9471	л [0 161]	[0 003]	0.402 (0.200)	-0.004 (0.306)
		locido	[0.347] R	[0.101]	[0.003]	(0.200)	(0.000)
		[0 77]	[0 012]				
		Relow	[0.012] R				
		[0 17]	[0 009]				
		[0.17]	[0.000]				

Table 9:	Tests	for	Market	Integration
	10010		mainer	megration

*Notes:* R and A indicate the null-hypotheses can be rejected and 'accepted' at the 5% level of significance. *p*-values and regime frequencies are given in square brackets. Standard errors of the estimated parameters and are shown in round brackets.

The finding of no thresholds for Ha Noi and Can Tho requires further comment, as these are the two most important demand centres in the Red River and Mekong River deltas. The absence of a threshold between these markets implies that there is a linear relationship between paddy prices in Ha Noi and

Can Tho. As can been seen from the time series plot of price differences on the right hand side of Figure 12, the inter-market prices spread only exceeds estimated transfer costs in 160 of our 161 observations. Under such circumstances one would not expect to observe a high correlation between paddy price changes in Ha Noi and Can Tho, and this is born out by the disparate cross plot of prices on the left-hand side of Figure 12. It is also reflected in the size of the coefficient on contemporaneous price changes ( $\gamma$ ) in Can Tho in the unrestricted linear error correction model being 0.410. Furthermore, the loading coefficient ( $\alpha$ ) on the lagged inter-market price spread indicates that the speed of price adjustment between Ha Noi and Can Tho is rather slow. Put differently, transfer costs between Ha Noi and Can Tho (which are 1,790 kms apart) are sufficiently high to make trade between the two markets unprofitable virtually all of the time. We can therefore conclude that paddy markets in Ha Noi and Can Tho are only weakly integrated which, in turn, suggests there is low integration between paddy and rice markets in the North and South of Vietnam.<sup>12</sup>

In contrast to Ha Noi and Can Tho, thresholds are indicated for all market pairs within the North and within the South of Vietnam. In general, we find two-threshold models, in accordance with the theory. However, for Ha Noi and Bac Ninh, the two geographically closest markets in our series we find no statistical significant difference between the parameters in the regime above the threshold (when the price in Ha Noi is greater than the price in Bac Ninh plus transfer costs) and the no-trade regime. This may reflect Bac Ninh's role as satellite town to Hanoi, with millers and wholesalers shipping to either Hanoi or Bac Ninh but little paddy or rice being traded between the two markets. Hence for this market pair a one threshold model is estimated, with strong co-movement of prices indicated by the  $\gamma$  coefficients (of 0.644 and 0.637) in the two regimes.

### Figure 12: Cross Plot of Ha Noi and Can Tho and Time Series Plot of the Inter-Market Price Spread with Thresholds Given by Estimated Transfer Costs



Source: Drawn by the authors using price data from the Information Center for Agriculture and Rural Development

<sup>&</sup>lt;sup>12</sup> Recall that the price series used for these tests is for ordinary paddy, so there may still be trade in high quality ('fancy') rice between the South and North of Vietnam.

Given the results about the number of thresholds we move on to test the hypothesis of long-run market integration for each regime in turn. Long-run integration, which corresponds to a one-unit change in the price in one market eventually translating into a one-unit price change in the other market, occurs in all regimes in all markets in the North and in all regimes in two of the three markets in the South. The exception is the market pair Can Tho – Ca Mau for which long run integration is rejected both in the no-trade regime and in the below regime (when the paddy price in Ca Mau exceeds the paddy price in Can Tho plus the transfer costs). Long run integration cannot, however be rejected in the upper regime, although the number of observations is very small in that regime, making us reluctant to draw strong conclusions for that regime.

Only when long-run market integration is indicated do we test for common dynamics and the Law of One Price using the restricted version of the threshold error correction model. Generally, for thresholds and market pairs for which long-run integration is indicated, common dynamics are also found.<sup>13</sup> This is not surprising because both cointegration and long-run integration imply the existence of a long-run relationship between paddy prices in the relevant markets. The finding of a common dynamics shows that all necessary information for predicting the future price is contained in the intermarket price spread (weak-form market efficiency).

The Law of One Price is, however, firmly rejected for most market pairs.<sup>14</sup> This is again not surprising as the form of the Law of One Price which is being tested requires both that price changes in the two markets move together on a one-for-one basis within a single period and that the common factor restriction also holds. As can be seen from the last but one column of Table 9, the  $\gamma$  coefficients between contemporaneous prices changes for these regimes and market pairs are all between 0.41 and 0.87, which is indicative of reasonably strong but far from perfect co-movement of paddy prices. Furthermore, for most market pairs, the  $\gamma$  coefficients exceed 0.6 when the inter-market price spread is above the upper or below the lower threshold. From this we may conclude that, when prices between markets are above or below their threshold levels, at least 60 percent of price changes are transmitted between markets within a month.

The final column of Table 9 shows the loading coefficients ( $\alpha$ ) from the restricted error correction models. As expected, all the  $\alpha$ s are negative showing that the inter-market price spreads will converge to their long-run equilibrium values. The absolute sizes of these loading coefficients indicate very rapid 'corrections' in some markets. In particular in the South we find that the inter-market price spreads returns to the no trade region within the following month. In addition, the absolute size of the loading coefficient is much smaller for Ha Noi and Can Tho than for market pairs within the North and within the South. This suggests that prices co-move more strongly when the threshold band is narrow. Put differently, when transfer costs do not prevent trade, spatial arbitrage brings price differentials back to their thresholds quickly. This is confirmed by the time series plots in Figure 13, which shows fairly rapid reversion of the price spreads whenever spreads exceed the transfer costs.

<sup>&</sup>lt;sup>13</sup> Here the exception is the lower regime in the Can Tho – An Giang market in which common dynamics is rejected with a p-value of 0.04.

<sup>&</sup>lt;sup>14</sup> When the Law of One Price is not rejected it is often in regimes with a relatively low frequency, such as the lower regime in the Ha Noi – Thai Binh market and the upper regime in the Can Tho – Tien Giang market. However, we find the Law of One Price to hold in two regimes in the Ha Noi – Nam Dinh market suggesting a close integration of this market pair.



Figure 13: Time Series Plots of the Inter-Market Price Spreads with Thresholds Given by Estimated Transfer Costs

Source: Drawn by the authors using price data from the Information Center for Agriculture and Rural Development

Taken together, our sequential tests for market integration show weak evidence of market integration between paddy markets in the North and South of Vietnam, but stronger evidence of spatial market integration within the North and within the South of Vietnam. While there is evidence of thresholds and long-run integration for most market pairs, the strict version of the Law of One Price does not hold in general even when the price difference between markets exceeds the

upper or lower thresholds. Nonetheless, when price spreads are above or below the relevant thresholds, a large fraction of the price changes are transmitted between paddy markets within a very short period of time. The extent and speed of price transmission is generally faster in paddy markets located in the South.

### 6. Conclusion and Policy Implications

There has been very little research on food market integration in Vietnam. This chapter aims to help fill this gap by focusing on the spatial integration of paddy markets in the Red River and Mekong River deltas. The empirical model developed uses estimates of transfer costs to generalize the well-known Ravallion model to allow for the possibility of threshold effects. A sequential testing strategy is then developed which progressively tests for market segmentation, the number of thresholds, long-run market integration, common dynamics and (the strict version of) the Law of One Price within an error-correction framework.

To determine whether an error correction framework is appropriate, we first test our eight monthly paddy price series for non-stationary and cointegration using the Elliot, Rothenberg and Stock version of the Dickey-Fuller test. All paddy price series are found to be non-stationary and cointegrated indicating that a long-run relationship exists between market pairs and that an error-correction formulation of our threshold model is appropriate.

When the unrestricted version of this model is estimated, we find weak evidence of market integration between paddy markets in the North and South of Vietnam and an absence of threshold effects. However, there is evidence of both threshold effects and stronger forms of spatial market integration for paddy markets within the North and within the South. Specifically, convergence towards the relevant thresholds and informational efficiency hold for most market pairs, with at least 60% percent of price changes being transmitted between markets within one month whenever price spreads exceeds their upper or lower thresholds. The extent and speed of price transmission within regional paddy markets is generally faster in the South than the North of Vietnam. However, the instantaneous version of the Law of One Price, which requires full price adjustment occur within a month, only holds for a few regimes and market pair.

The policy implications of these results flow from the answers to our three initial questions. First, since there is limited evidence of integration between paddy markets in the North and South of Vietnam, national level policies cannot be relied upon to stabilize or support paddy prices. Instead, agricultural policies need to be designed and implemented with the specific production, consumption and marketing characteristics of northern and southern Vietnam in mind. Second, since there is evidence of market integration within the Red River and Mekong River deltas, paddy markets within these regions can be relied upon to transmit price signals between deficit and surplus areas relatively well. This is good news as it should help agricultural producers to specialize according to comparative advantage and promote investment and growth in the rural areas. Third, since the speed and extent of price transmission is relatively rapid within the North and within the South of Vietnam, the private sector trade can be relied upon to transfer rice and paddy between markets in an efficient manner. If, however, large demand-supply imbalances were to emerge between the North and South, transfer costs are likely to prevent private sector trade taking place and the public sector might need to intervene to ensure adequate supplies in the short-term. Such interventions would, however, need to be implemented in a consistent and market friendly way to avoid destabilizing and distorting an essentially well integrated and competitive paddy marketing system.

Finally further research is needed to examine the robustness of our results to different assumptions about threshold costs and investigate where and how paddy and rice price formation takes places (including the influence of the rice export market). There is also a need to collect better and more frequent data on paddy and rice prices in different locations and stages in the marketing chain, so that the vertical dimensions of market integration can be examined.

### References

- Alexander, C., and J. Wyeth. 1994. "Cointegration and market integration: an application to the Indonesian rice market." *Journal of Development Studies* 30: 303-28.
- Baulch, B. 1997. "Transfer costs, spatial arbitrage, and testing for food market integration." *American Journal of Agricultural Economics* 79: 477-87.
- Baum, C. F., and R Sperling. 2001. "DFGLS: Stata module to compute Dickey-Fuller/GLS unit root tests." Statistical Software Components S410001. Boston College Department of Economics.
- Benjamin, D., and L. Brandt. 2004. "Agriculture and income distribution in rural Vietnam under economic reforms: a tale of two regions." In Glewwe, P., N. Agrawal, and D. Dollar eds., *Economic Growth, Poverty and Household Welfare in Vietnam*, World Bank Regional and Sectoral Studies, Washington, DC.
- Dawson, P. J., and P. Dey. 2002. "Testing for the Law of One Price: rice market integration in Bangladesh." *Journal of International Development* 14: 473-84.
- Elliott, G., T. J. Rothenberg , and J. Stock. 1996. "Efficient tests for an autoregressive unit root." *Econometrica* 64: 813-36.
- FAOSTAT. "FAOSTAT Data 2006". http://faostat.fao.org/faostat/collections?subset=agriculture, last accessed February 2006.
- Goodwin, B. K., and N. E. Piggott. 2001. "Spatial market integration in the presence of threshold effects." *American Journal of Agricultural Economics* 83(2): 302-17.
- GSO [General Statistical Office]. 2006. Statistical Yearbook of Vietnam 2005. Statistical Publishing House.
- Hansen, B. E. 1996. "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica* 64: 413–430.
- Hansen, B. E. 1997. "Inference in TAR models." Studies in Nonlinear Dynamics and Econometrics 2: 1-14.
- Hansen, B. E., and B. Seo. 2002, "Testing for two-regime threshold cointegration in vector error correction models." *Journal of Econometrics* 110: 293-318.
- Isard, P. 1977. "How far can we push the "Law of One Price?"." American Economic Review 67(5): 942-48.
- Jones, W. O. 1968. "The structure of staple food marketing in Nigeria as revealed by price analysis." Food Research Institute Studies 8(2): 95-124.
- Lele, U. J. 1967. "Market integration: a study of sorghum prices in Western India." *Journal of Farm Economics* 49(1): 149-59.
- Luu, Thanh Duc Hai. 2003. "The organization of the liberalized rice market in Vietnam." PhD thesis. University of Groningen, Netherlands.
- Meyer, J. 2004. "Measuring market integration in the presence of transaction costs—a threshold vector error correction approach." *Agricultural Economics* 31: 327-34.
- Minot, N., and F. Goletti. 1999, "Rice market liberalization and poverty in Vietnam." IFPRI Research Report 114. International Food Policy Research Institute, Washington, D.C.
- Ravallion, M. 1986. "Testing market integration." American Journal of Agricultural Economics 68: 102-09.
- Richardson, J. D. 1978. "Some empirical evidence on commodity arbitrage and the Law of One Price." Journal of International Economics 8: 341-51.
- Sarno, Lucio, M. P. Taylor, and I. Chowdhury. 2004. "Nonlinear dynamics in the deviations from the Law of One Price: a broad-based empirical study." *Journal of International Money and Finance* 23: 1-25.
- Sexton, R. J., C. L. Kling, and H. F. Carman. 1991. "Market Integration, Efficiency of Arbitrage and Imperfect Competition: Methodology and an Application to U.S. Celery." American Journal of Agricultural Economics 17(2): 568-80.

## 2

### **Regional Labor Market Integration**

Hoang Thanh Huong, Le Van Chon, Le Dang Trung and Remco Oostendorp

This chapter analyses the regional integration of labor markets in Vietnam. Using evidence from 4 waves of household surveys over the period 1993-2004, the paper finds that regional wage levels have been diverging during the period 1993-2004, although there is convergence in the period 2002-2004. Regional wage gaps can increasingly be explained by regional differences in human capital, industrial structure and ownership, but even after controlling for (changes in) these factors, hourly wages are still diverging across regions in Vietnam. However, we do find that regional wage gaps with *neighboring* regions are converging after we control for regional differences in human capital, industrial structure and ownership, suggesting that labor markets are *locally converging but globally diverging* in Vietnam.

The chapter also estimates regional shadow wages in agricultural self-employment, and finds that shadow wages are significantly lower than market wages, confirming a lack of integration between wage- and self-employment and the existence of surplus labor in rural areas. However, shadow wages as a ratio of market wages have increased between 2002 and 2004 for the whole country, suggesting that the markets for wage and self-employment are becoming more integrated in Vietnam. At the same time, the integration is found to be the weakest in the North East and the North West, and the strongest in the South East and the Mekong River Delta.

In summary, for Vietnam as a whole, we conclude that regional labor markets in Vietnam have been diverging, but there are signs of increasing local integration.

### 1. Introduction

There is increasing evidence that the benefits of market integration are unevenly shared across households and regions in Vietnam. Aggregate economic performance has been remarkable and Vietnam is now seen as one of the most successful economies of transition having experienced a growth rate of 7.4 percent annually on average over the period 1991-2003. At the same time there is a steady tendency towards greater inequality of per capita expenditures, albeit at a moderate pace. Also regional differences in poverty incidence are considerable, with a poverty rate of 68% in the North West against only 10.6% in the South East in 2002, and with poverty increasingly concentrated in the poor North Central Coast and the Central Highlands (Vietnam Development Report 2004, tables 1.2-1.5).

Similar tendencies have been observed for the labor market in Vietnam. Wage employment has

been steadily increasing since 1993 (Vietnam Development Report 2006, table 7.1) but the benefits are unevenly shared as job creation takes place predominantly in urban areas and industrial areas. Also labor markets remain *regionally segmented* in the sense that workers with identical levels of human capital face different rates of return depending on where they work reinforcing the urban-rural income gap (Gallup 2002, ADB 2005).

In this chapter we will analyze the regional integration of labor markets in Vietnam. In particular we will address three questions. First, we ask whether there is regional convergence or divergence of wages in Vietnam. Although earlier studies have reported significant wage differences across regions and the existence of labor market segmentation, they have not looked at the question of regional wage convergence specifically. We find that regional wage levels have been diverging during the period 1993-2004, although there is some convergence between 2002 and 2004.

Second, we ask how much of the regional variation in wages can be explained by inter-regional differences in human capital, industrial structure and/or ownership, and whether the observed divergence in hourly wages across regions can be explained by regional changes in these factors. The analysis shows that regional wage gaps can increasingly be explained by regional differences in human capital, industrial structure and ownership. Also it is found that even after controlling for changes in these factors, hourly wages are still diverging across regions in Vietnam. However, it is also found that regional wage gaps with neighboring regions are converging, suggesting that there is local convergence but global divergence.

Third, we also ask to what extent the labor markets for wage- and self-employment are integrated. Most workers are self-employed and do not earn market wages but shadow wages. In principle, if the labor markets for wage- and self-employment are fully integrated, market and shadow wages should be equal, and an analysis of market wages suffices. If these markets are segmented, for instance, because of barriers to job mobility, lack of job market information, and discrimination, then market and shadow wages will differ. We therefore also estimate regional shadow wages to provide a more comprehensive picture of labor market integration in Vietnam. We do find that shadow wages are significantly lower than market wages, confirming a lack of integration between wage- and self-employment and the existence of surplus labor in rural areas. However, shadow wages as a ratio of market sfor wage and self-employment are increasingly becoming integrated in Vietnam. However, the integration is the strongest in the South East and the Mekong River Delta, and the weakest in the North East, suggesting that the integration of markets for wage and self-employment is local.

In section 2 of this chapter we first analyze a number of descriptive features of regional macro and micro convergence in Vietnam. Most studies of convergence focus on macro variables, such as GDP per capita, with less attention being paid to micro variables, such as wages and prices. In section 3 we will analyze whether regional wage differences can be explained by regional differences in human capital and/or industrial structure. In section 4 we turn to the estimation and analysis of regional differences in shadow wages. Section 5 concludes.

### 2. Macro Versus Micro Convergence

Rassekh and Thompson (1998) distinguish between two categories of convergence, namely *micro* convergence and *macro* convergence. Micro convergence refers to a tendency towards the equalization of factor returns, such as wages, rents and prices, across geographical entities such as countries, regions, provinces, and districts ('economies'). The factor-price equalization theorem of the Heckscher-Ohlin-Samuelson model of international trade predicts such micro-convergence of prices, wages and rents under conditions of free trade, identical technology but different factor endowments.

Macro convergence refers to the tendency of aggregate variables such as per capita income or output per worker to become more similar across economies across time. The neoclassical growth model developed by Solow (1956), for instance, suggests that the steady-state per capita income levels are independent from the initial state and therefore poor economies will converge towards richer economies as long as they have identical steady-states. Other theories, however, have focused on the possibility of divergence as well, for instance due to threshold externalities, capital market imperfection, heterogeneity, country size and club formation (see references in Quah 1996).

There are two reasons to study both macro and micro convergence. First, macro and micro convergence are related but not identical concepts. Since per capita income is a weighted average of factor prices, macro convergence will follow from micro convergence if per capita endowments are similar, but micro convergence can create macro divergence otherwise (Rassekh and Thompson 1998). Second, macro and micro convergence are important for different reasons. Macro convergence is primarily important in the debate on competitiveness and inequality, while micro convergence is important in the debate on market functioning and imperfections.

In the economics literature a number of different concepts have been developed to measure macro and micro convergence, of which  $\sigma$ - and  $\beta$ -convergence are the most well known and frequently used.<sup>15</sup> The concept of  $\sigma$ -convergence implies that the standard deviation of the variable of interest (e.g. log per capita income, log hourly wages) across the different economies tends to decrease over time.<sup>16</sup> The concept of  $\beta$ -convergence on the other hand says that there is  $\beta$ -convergence if poor economies (e.g. low log per capita income, low hourly wages) tend to grow faster than rich ones.<sup>17</sup> These concepts are closely related and it can be shown that  $\beta$ -convergence is a necessary but not sufficient condition for  $\sigma$ -convergence (Sala-i-Martin 1996). We therefore will use the concept of  $\sigma$ convergence to measure convergence in this paper.

There are only very few studies on Vietnam applying formal convergence analysis. One of the studies is Klump and Anh (2004) in which the authors have applied  $\sigma$ - and  $\beta$ -convergence tests to gross regional product per worker in Vietnam's 61 provinces over the period 1995-2000. They find that there is limited  $\sigma$ -convergence between 1995 and 2000 for gross regional product per worker (the  $\sigma$ 

<sup>16</sup>  $\sigma_{t+T} < \sigma_t$  for T>0.

<sup>17</sup> This is the definition for absolute  $\beta$ -convergence as opposed to conditional  $\beta$ -convergence. Let  $y_{i,t}$  be the variable of interest for an economy *i* in period *t* and  $\gamma_{i,t,t+T} \equiv \log(y_{i,t+T} / y_{i,T})/T$  is the economy's *i*'s annualized growth rate of  $y_{i,t}$  at time *t*. If the regression  $\gamma_{i,t,t+T} = \alpha - \beta \log(y_{i,t}) + \varepsilon_{i,t}$  is estimated and  $\beta$ >0 then it is said that the economies exhibit absolute  $\beta$ -convergence. If the regression also includes a number of control variables because different economies might converge to different steady states, then if  $\beta$ >0 it is said that the economies exhibit convergence (Sala-i-Martin 1996).

<sup>&</sup>lt;sup>15</sup> There are two other important concepts of convergence, namely stochastic convergence and intradistributional dynamics. Stochastic convergence implies that the difference in the variable of interest across economies is stationary around zero (no unit roots) (Bernard and Durlauf 1996). Unlike the cross-section tests of  $\beta$ - and  $\sigma$ -convergence, this is a time-series test. However it may have poor power properties when applied to data from economies in transition because it assumes that sample moments are an appropriate proxy for asymptotic means (idem, p.171). Intra-distributional dynamics analyzes the evolution of entire distributions and is more informative than  $\beta$ - and  $\sigma$ -convergence because it can also reveal tendencies for clustering (polarization) and the mobility between different groups of economies (Quah 1996). However the application of intra-distributional dynamics requires panel information on a large number of economies and is therefore less appropriate for our regional analysis of 10 regions.

falls from 0.57 to 0.55). They also find (absolute)  $\beta$ -convergence, but the convergence rate is estimated at a low 1.4% implying a half-life of 50 years.<sup>18</sup>

The existing convergence studies therefore suggest that there is macro convergence in Vietnam but it remains unclear whether the same conclusion holds with other indicators as well as for longer time periods. We therefore look at two other indicators, namely per capita expenditures and hourly wages. Per capita expenditures are an important indicator of macroeconomic performance and are widely used as a proxy for economic welfare and permanent income. Hourly wages measure the returns to labor on the labor market and are therefore a crucial indicator of microeconomic performance.

We use the Vietnam (Household) Living Standard Surveys to measure per capita expenditures and hourly wages at the regional level in Vietnam for 1993, 1998, 2002 and 2004. While the 2002 and 2004 surveys are representative at the provincial level, the 1993 and 1998 surveys are only representative at the regional level, and therefore the surveys allow convergence analysis at the regional level but not provincial level.<sup>19</sup>

In Figure 1 we report the regional levels of per capita expenditures. Per capita expenditures are in current prices but deflated by regional and monthly prices.<sup>20</sup> Also the per capita expenditures are weighted by the sampling weights to be representative of the population averages. We distinguish among 10 regions, namely Red River Delta (excluding Hanoi), North East, North West, North central Coast, South Central Coast, Central Highlands, South East (excluding Ho Chi Minh City), Mekong River Delta, Hanoi and Ho Chi Minh City. We order the regions from poorest to richest for each year and we normalized the per capita expenditures of each region by the per capita expenditures of the poorest one.

The figure suggests that regional per capita expenditures have been diverging across regions between 1993 and 2004, although there is some convergence between 2002 and 2004. The standard deviation of regional log per capita expenditures increased from 0.32 in 1993, to 0.37 in 1998, 0.47 in 2002 and declined to 0.43 in 2004. Hence, we conclude that there is  $\sigma$ -divergence in regional per capita expenditures between 1993 and 2002,  $\sigma$ -convergence between 2002 and 2004, and  $\sigma$ -divergence over the entire 1993-2004 period. Similar conclusions hold if we calculate *weighted* standard deviations of regional log per capita expenditures, with weights equal to regional population sizes. In this case the standard deviation increased from 0.27 in 1993, to 0.30 in 1998, 0.34 in 2002 and declined to 0.33 in 2004.

<sup>&</sup>lt;sup>18</sup> The  $\beta$  coefficient can be interpreted as a function of the *rate of convergence* (the percentage of the gap between the steady state level and current level that is closed each period):  $\beta = (1 - e^{-\gamma T})/T$ , where  $\gamma$  is the rate of convergence. For the derivation see Rassekh (1998, pp.90-91).

<sup>&</sup>lt;sup>19</sup> Because the sampling was done by stratified random sampling across communes and wards throughout the country, the 1993 and 1998 surveys are also in principle 'representative' at the provincial level. However, given the small sample sizes of the 1993 and 1998 survey, the samples at the province-level are too small to provide reliable estimates for 1993 and 1998.

<sup>&</sup>lt;sup>20</sup> Monthly price deflators are necessary because the survey was executed across several months.



Figure 1: Per Capita Expenditures across Regions in Vietnam, 1993-2004

An interesting question is whether the observed macro divergence in per capita expenditures is matched by micro divergence in regional wage levels. It was already noted that regional micro and macro convergence are related but not identical, and therefore it is an open question whether the observed macro divergence is mirrored by micro divergence in regional labor markets. In the next figure we therefore redo the same convergence analysis but now for hourly wages. Hourly wages are calculated for all respondents between ages 15 and 65 who report wage employment as their main activity in the past 12 months and are deflated by regional and monthly price deflators and sampling weights have been applied to calculate regional averages. Because the number of individuals with wage employment was very low for the regions North West and Central Highlands in 1993 and 1998, these regions are excluded for 1993 and 1998.

Figure 2 shows a very similar pattern of convergence as Figure 1. Once again, hourly wages have been standardized by the hourly wage in the lowest wage region and regions have been ranked by hourly wage level. Because the regions Central Highlands and North West were respectively the 4<sup>th</sup> and 6<sup>th</sup> lowest wage regions in 2002 and 2004, we left these regions empty in the figure for 1993 and 1998. Wages in the Red River Delta were among the lowest in the country, while wages in Ha Noi and HCMC were the highest. Regional log hourly wages have been diverging across regions between 1993 and 2004, although there is some convergence between 2002 and 2004. The standard deviation of regional log per capita expenditures increased from 0.15-0.16 in 1993-1998, to 0.23 in 2002 and declined to 0.20 in 2004.<sup>21</sup> Similar conclusions hold if we calculate *weighted* standard deviations of regional log hourly wages, with weights equal to regional population sizes. In this case the standard deviation

<sup>(</sup>Poorest Region Normalized at One)

Source: Own calculations based on 1992/93 and 1997/98 VLSSs and 2002 and 2004 VHLSSs

<sup>&</sup>lt;sup>21</sup> Excluding the regions North West and Central Highlands. However, the standard deviation of the log hourly wage is quite similar if we include these regions in the calculations for 2002 and 2004 as well: 0.21 for 2002 and 0.18 for 2004.

increased from 0.13-0.15 in 1993-1998, to 0.20 in 2002 and declined to 0.17 in 2004.<sup>22</sup> Hence, we conclude that there is  $\sigma$ -divergence in regional log hourly wages between 1993 and 2002,  $\sigma$ -convergence between 2002 and 2004, and  $\sigma$ -divergence over the entire 1993-2004 period.





Source: Own calculations based on 1992/93 and 1997/98 VLSSs and 2002 and 2004 VHLSSs

<sup>&</sup>lt;sup>22</sup> Including the regions North West and Central Highlands the standard deviations are 0.19 and 0.17 in 2002 and 2004 respectively.



Figure 3: Unexplained Logarithmic Hourly Wage Gaps across Regions in Vietnam, 1998-2004

Source: Own calculations based on 1992/93 and 1997/98 VLSSs and 2002 and 2004 VHLSSs

These findings suggest that regional labor markets have become less integrated between 1993 and 2004. However, the analysis only looks at *average* hourly wages, and in the next section we will analyze whether our conclusion of wage divergence needs to be modified if one takes into account the heterogeneity of labor.

Before turning to the next section, however, we address one more issue. The main focus of this paper is *between*-region variation in hourly wages rather than *within*-region variation. This focus is justified as we are interested in regional market integration, assuming that labor markets are spatially (regionally) disintegrated. Also the Vietnam (Household) Living Standard Surveys do not allow convergence analysis at a lower level of aggregation. It is important to note, however, that most of the variation in hourly wages is within-region rather than between-region. Simple analysis of variance shows that in the period 1993-2004 only between 4.4 and 10.4% of the total variance in hourly wages can be explained by between-regional variation, leaving the remainder to within-region variation.<sup>23</sup> This may seem very low but we need to consider two points. First, the contribution of between-regional variation is probably severely underestimated because an important part of the total variation may be simply measurement error in the hourly wage variable. And second, the contribution of between-regional variation forms 28.8-40.6% of the explained variation in a standard Mincer wage regression.<sup>24</sup> Therefore, regional wage variation is an important component of total wage variation even if it can not explain all the existing variation.

<sup>&</sup>lt;sup>23</sup> After controlling for human capital differences (experience and years of schooling) as well as for gender, we find that between 3.2 and 13.2% of the total variance is due to between-region variation.

<sup>&</sup>lt;sup>24</sup> The R<sup>2</sup> is between 0.16 and 0.36 in the Mincer equation (see Table 1).

### 3. Explaining Regional Wage Differentials

In the previous section we found large differences in regional wage differences and also that these wage differences diverged between 1993 and 2004 but converged between 2002 and 2004. In this section we ask the following questions. First, how much of the regional variation in wages can be explained by inter-regional differences in human capital and/or industrial structure? And second, if we control for inter-regional differences in human capital, industrial structure and/or ownership, how does this modify our wage convergence analysis? In other words, does the observed wage divergence also occur if we control for worker and employment heterogeneity?

These questions are important to understand whether the observed interregional wage differences are labor market imperfections leading to different returns to human capital across regions or simply reflecting interregional differences in human capital. If it is the former, then regional development policies should focus on labor market development. If it is the latter, then regional development should focus on additional investments in human capital to improve wage levels in the low wage regions.

#### Decomposition analysis of regional wage gaps

In order to answer the question how much of the regional differences in wages can be explained by differences in human capital, industrial structure and/or ownership, we apply the Blinder-Oaxaca decomposition method. Let  $w_{ir}$  be the log hourly wage of worker *i* in region *r*. The first step in the Blinder-Oaxaca decomposition is to estimate a Mincer wage equation for workers *i* in each region:

$$w_{ir} = X_{ir}\beta_r + \varepsilon_{ir} \qquad \text{for } \forall r \in \{1, \dots, R\}$$
(1)

where  $X_{ir}$  is a vector of wage-determining factors,  $\beta_r$  a vector of coefficients,  $\varepsilon_{ir}$  an error term with mean zero, and R is the total number of regions. If Equation (1) is estimated by ordinary least squares for each  $r \in \{1, ..., R\}$ , then the mean regional wage  $\overline{w}_r$  can be written as:

$$\overline{w}_r = X_r \beta_r \tag{2}$$

where  $\overline{X}_r$  is the regional mean of  $X_{ir}$  and  $\hat{\beta}_r$  the estimated  $\beta_r$ . The Blinder-Oaxaca decomposition of the wage gap between region *r* and region *s* is then given by:

$$\overline{w}_r - \overline{w}_s = (\overline{X}_r - \overline{X}_s)\beta_s + \overline{X}_s(\beta_r - \beta_s) + (\overline{X}_r - \overline{X}_s)(\beta_r - \beta_s) = E + C + CE$$
(3)

Depending on the model that is assumed to be the "true" model, the three terms of the decomposition may be used to determine the "explained" (M) and "unexplained" (U) parts of the differential. Oaxaca (1973) proposed assuming either the model for region *s* or region *r* as the no-discrimination model, which implies that either M=E and U=C+CE or M=E+CE and U=C, respectively. More generally, the coefficients of the "true" model may be expressed as

$$\beta^* = W\beta_r + (I - W)\beta_s \tag{4}$$

where I is an identity matrix and W is a matrix of weights. Analogously, the decomposition may be written as

$$\overline{w}_r - \overline{w}_s = (\overline{X}_r - \overline{X}_s)(W\beta_r + (I - W)\beta_s) + [(I - W)\overline{X}_r + W\overline{X}_s](\beta_r - \beta_s)$$
(5)

In the two cases proposed by Oaxaca (1973), W is a null matrix or equals I, respectively. Alternatively, Neumark (1988) proposed using the coefficients from a pooled model for both groups, which implies that

$$\overline{w}_r - \overline{w}_s = (\overline{X}_r - \overline{X}_s)\beta^* + [\overline{X}_r(\beta_r - \beta^*) + \overline{X}_s(\beta^* - \beta_s)]$$
(6)

where  $\beta^*$  is the vector of the coefficients from the pooled model. In this paper we follow Neumark's approach as it lies in between the two extremes proposed by Oaxaca originally.

We apply the Blinder-Oaxaca-Neumark decomposition method to decompose the wage gaps

between any region and its *neighboring* regions. We make the comparison with neighboring regions because we expect labor market integration to be the strongest between spatially proximate regions.<sup>25</sup> The Blinder-Oaxaca decomposition will tell us how much of these neighboring regional wage gaps can be explained by differences in endowments and how much is left unexplained. Given that the decomposition method is applied to each of the 1993, 1998, 2002 and 2004 surveys, it will therefore be possible to test whether the relative importance of endowments for understanding regional wage gaps has increased over time.

However, it is important to note that even if this is the case, this does not necessarily imply that labor markets are increasingly integrated. First of all, the role of endowments in explaining regional wage gaps may increase simply because the regional distribution of endowments increasingly favors high wage regions. For instance, if the better educated tend to migrate to the high wage areas, then regional differences in endowments will increasingly be able to explain (increasing) regional wage differences. In this case, an increasing role for endowments reflects changes in the regional distribution of endowments.

Second, the role of endowments in explaining regional variation in wages may also increase because wages increasingly reflect endowments and become less a function of idiosyncratic factors. For instance, if wages in Vietnam are increasingly market-determined and reflecting differences in human capital, then we expect that regional differences do increasingly reflect regional differences in human capital as well. In this case an increasing role for endowments in explaining regional wage differences reflects improved labor market functioning.

Table 1 reports the Mincer regressions for 1993, 1998, 2002 and 2004. The dependent variable is the logarithm of hourly wages from the main job for individuals between 15 and 65 years. Hourly wages are corrected for regional price differences. The regressions control for years of education and experience (measured as age minus years of schooling, gender (dummy for female), industry (with agriculture as the reference group), ownership of employer (with private firms as the reference group), and regions (with North Central Coast as the reference group). An interaction term is included for years of education and gender because earlier studies have found that the returns to education are different between males and females in Vietnam (Nguyen *et al.* 2006,). Also a Heckman sample selectivity correction term is included to control for the fact that the regression is estimated for individuals with wage employment only.<sup>26</sup> Appendix A provides the descriptive statistics of the model variables.

<sup>&</sup>lt;sup>25</sup> This assumption is confirmed in our analysis. For instance for 2004 we find that 59.4% of the wage gap between neighboring regions can be explained by regional differences in human capital, industrial structure and ownership, (table 2) against only 49.4 for the wage gap between any region and all the other regions (not reported).

<sup>&</sup>lt;sup>26</sup> The participation regression includes the following variables: age, years of education, gender, the share of children in the household, the household size and the amount of land owned by the household.

Dependent variable: log hourly wages	1993	1998	2002	2004
Female	-0.23	-0.17	-0.23	-0.30
	(0.01)	(0.00)	(0.00)	(0.00)
Years of schooling	-0.01	0.03	0.04	0.04
	(0.59)	(0.00)	(0.00)	(0.00)
Female x years of schooling	0.02	0.00	0.01	0.02
	(0.10)	(0.92)	(0.00)	(0.00)
Experience	0.01	0.00	0.03	0.03
	(0.18)	(0.61)	(0.00)	(0.00)
Experience squared (10-3)	-0.07	0.04	-0.42	-0.44
	(0.60)	(0.68)	(0.00)	(0.00)
Mining	-0.11	0.21	0.24	0.27
	(0.34)	(0.02)	(0.00)	(0.00)
Manufacturing	0.02	-0.09	0.00	-0.04
	(0.70)	(0.04)	(0.85)	(0.05)
Electricity, construction	0.24	0.20	0.11	0.09
	(0.00)	(0.00)	(0.00)	(0.00)
Commerce	-0.15	-0.09	0.01	0.00
	(0.09)	(0.12)	(0.60)	(0.96)
Transportation, communication	0.06	0.15	0.21	0.22
	(0.49)	(0.02)	(0.00)	(0.00)
Finance, other services	-0.25	-0.31	0.04	-0.16
	(0.00)	(0.01)	(0.02)	(0.00)
Government	-0.20	-0.25	0.11	0.38
	(0.00)	(0.00)	(0.00)	(0.00)
SOE	-0.15	-0.08	0.19	0.10
	(0.00)	(0.03)	(0.00)	(0.00)
FDI	0.15	0.11	0.28	0.29
	(0.52)	(0.01)	(0.00)	(0.00)
Inverse Mill's Ratio	-0.34	0.02	-0.03	-0.03
	(0.00)	(0.60)	(0.31)	(0.41)
Constant	1.01	1.33	0.40	0.59
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	1863	2244	19194	5778
R2	0.16	0.13	0.31	0.36

Table 1: Mir	ncer Rearess	sions 1993-2004
--------------	--------------	-----------------

*Note*: Robust p values in parentheses. The regressions also include regional dummies.

Source: Own calculations based on 1992/93 and 1997/98 VLSSs and 2002 and 2004 VHLSSs

The results of Table 1 show a number of interesting findings. First the returns to education have increased during the period 1993 and 2004. Second, females earn significantly less than males, but the gender gap is declining with education. Third, there are large unexplained industry-wage differentials. Fourth, workers in the private sector tend to be paid less than similar workers (in terms of human capital) in Government, SOEs and FDIs in 2002 and 2004. In 1993 and 1998, however, Government and SOE workers tended to be paid less. This reversal in Government and SOE pay may reflect the outcome of the restructuring efforts of SOEs and of the salary increase policy of the Government. Since 2002, the number of workers in SOES has been reduced substantially due to the

reorganization of state enterprises<sup>27</sup>. In the Government service sector, however, there has always been high pressure for the Government to increase salaries for their staff because servants always complain that the salaries they receive are far not enough for their families. Accordingly, the Government of Vietnam has implemented a wage reform for the period 2003-2007. Since 2004, wage increase has been taken place two times.

For the decomposition analysis we run the Mincer regressions separately for each region and its neighboring region(s). In the case there are multiple neighboring regions, regional dummies are included in the regression for the neighboring regions because they might be correlated with the other regressors and their omission would lead to omitted variable bias. However, in the decomposition analysis we look at the *average* neighboring region and analyze how much of the regional wage gap can be explained by human capital, industry, and ownership variables.<sup>28</sup> Table 2 reports the results from the decomposition analysis where we take into account the sampling weights. For 1993 and 1998 we omit the regions North West and Central Highlands because there are too few observations for wage earners available in those years. The last column summarizes the results for all region by calculating how much of the aggregate wage difference can be explained.<sup>29</sup>

Table 2 shows a number of interesting results. First, hourly wages are consistently low in the Red River Delta relative to its neighboring regions (North East, North West, North Central Coast and Hanoi) over the period 1993-2004. Second, hourly wages are relatively high in Hanoi and HCMC compared to the neighboring regions. Third, the same holds for the unexplained wage gap between the Red River Delta, Hanoi and HCMC and its neighboring regions, suggesting that there is a strong incentive to migrate from and to these regions. Fourth, the wage gap with neighboring regions has significantly narrowed for the North East, North Central Coast, and South Central Coast over the period 1993-2004, suggesting increasing (local) integration. Fifth, while only 18.7% of the average wage gap could be explained in 1993, this increased to 26.9% in 1998, and 59.0 respectively 59.4% in 2002 and 2004.<sup>30</sup> Hence an increase part of the regional wage gaps can be explained by regional differences in human capital, industrial structure and ownership.

However, as noted before, the increasing role of endowments may reflect changes in the regional distribution of endowments and/or improved labor market functioning. However it does not imply that labor markets are becoming increasingly regionally integrated in Vietnam as this depends on dispersion in the unexplained regional wage gaps.

<sup>&</sup>lt;sup>27</sup> The policy is stated in Decree No. 41/2002/ND-CP of the Government.

<sup>&</sup>lt;sup>28</sup> The average neighboring region is calculated as having an average value of each of the regional dummy coefficients.

<sup>&</sup>lt;sup>29</sup> This is calculated as [ $\Sigma$  Explained \* sign(Difference)]/ [ $\Sigma$  |Difference|].

<sup>&</sup>lt;sup>30</sup> Excluding the North West and Central Highlands the percentages are 58.9 and 59.0% in 2002 respectively 2004.

				NOO	000		05	MDD			Average
	RKD	NE	NW	NCC	SCC	СН	SE	MRD	HN	НСМ	(%)
1993											
Difference	-0.43	-0.31	NA	-0.53	-0.29	NA	-0.07	-0.22	-0.08	0.20	
	(0.10)	(0.11)	NA	(0.09)	(0.09)	NA	(0.12)	(0.09)	(0.09)	(0.07)	
Explained	-0.03	-0.10	NA	-0.24	0.07	NA	0.02	-0.10	0.13	0.15	18.7
	(0.03)	(0.04)	NA	(0.05)	(0.05)	NA	(0.04)	(0.05)	(0.06)	(0.05)	
Unexplained	-0.40	-0.21	NA	-0.28	-0.36	NA	-0.09	-0.12	-0.21	0.05	81.3
	(0.10)	(0.10)	NA	(0.09)	(0.07)	NA	(0.15)	(0.11)	(0.06)	(0.03)	
1998											
Difference	-0.26	-0.10	NA	0.25	-0.02	NA	-0.12	-0.01	0.25	0.20	
	(0.10)	(0.09)	NA	(0.07)	(0.05)	NA	(0.05)	(0.05)	(0.09)	(0.05)	
Explained	0.04	-0.08	NA	0.02	0.08	NA	-0.03	-0.06	0.08	0.18	26.9
-	(0.03)	(0.04)	NA	(0.04)	(0.02)	NA	(0.02)	(0.02)	(0.08)	(0.03)	
Unexplained	-0.30	-0.02	NA	0.23	-0.09	NA	-0.09	0.05	0.17	0.02	73.1
	(0.10)	(0.09)	NA	(0.06)	(0.05)	NA	(0.05)	(0.04)	(0.08)	(0.03)	
2002											
Difference	-0.22	0.01	0.06	0.01	0.09	-0.13	-0.03	-0.38	0.49	0.54	
	(0.03)	(0.04)	(0.11)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.05)	(0.04)	
Explained	-0.14	0.06	0.12	0.03	0.03	0.01	0.02	-0.22	0.33	0.24	59.0
	(0.02)	(0.03)	(0.05)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.06)	(0.04)	
Unexplained	-0.08	-0.05	-0.06	-0.02	0.06	-0.14	-0.05	-0.16	0.16	0.30	41.0
·	(0.02)	(0.03)	(0.08)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.06)	(0.03)	
2004	. ,	. ,	. ,	. ,	. ,	. ,	. ,	. ,	. ,	. ,	
Difference	-0.21	0.00	0.13	-0.04	0.04	-0.06	0.00	-0.32	0.47	0.36	
	(0.04)	(0.05)	(0.08)	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)	
Explained	-0.13	0.08	0.13	0.03	0.01	0.00	0.02	-0.17	0.30	0.17	59.4
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	
Unexplained	-0.08	-0.08	0.01	-0.07	0.03	-0.07	-0.02	-0.15	0.17	0.19	40.6
	(0.03)	(0.05)	(0.07)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	

Table 2: Regional Decomposition Analysis 1993-2004 (Standard Errors in Parentheses)

Note: in 1993 and 1998 there are too few wage observations for the regions North West and Central Highlands to decompose the wage gap. Source: Own calculations based on 1992/93 and 1997/98 VLSSs and 2002 and 2004 VHLSSs.

### Convergence analysis after controlling for regional differences

The analysis provides us with two measures of the unexplained regional wage gaps. First, we have the estimated coefficients of the regional dummy variables in the Mincer regressions (regional dummies were included in the regressions but the coefficients were not reported in Table 1). Second, we have the unexplained wage gaps in the Blinder-Oaxaca decomposition (Table 2).

The standard deviation of the regional dummy coefficients shows a pattern similar to that for the log hourly wages: the standard deviation increased from 0.13-0.18 in 1993-1998, to 0.23 in 2002 and declined to 0.19 in 2004.<sup>31</sup> Hence, even after controlling for regional differences and changes in human capital, industrial structure and ownership, we still find divergence of regional wages over the period 1993-2004, although with some convergence in 2002-2004.

<sup>&</sup>lt;sup>31</sup> Excluding the regions North West and Central Highlands.

However, the pattern of the unexplained wage gap from the decomposition analysis shows *convergence* over the period 1993-2004 (Figure 3).<sup>32</sup> The lowest unexplained wage gap is normalized at zero and the unexplained wage gaps are ranked in increasing order in figure 3. The standard deviation of the unexplained wage gaps did first increase from 0.15 in 1993 to 0.17 in 1998, but did subsequently fall to 0.15 in 2002 and 0.12 in 2004. This suggests that labor markets are increasingly becoming integrated in Vietnam.

How can we explain the different results for our two measures of unexplained wage gaps –regional dummy coefficients from the Mincer regressions and unexplained wage gaps from the Blinder-Oaxaca decomposition? The answer is that the unexplained wage gaps from the Blinder-Oaxaca decomposition are with respect to the *neighboring* regions rather than with respect to all other regions. If we do the decomposition of the wage gap with respect to all other regions (not reported), we find virtually the same pattern of divergence as for the regional dummy coefficients from the Mincer regressions: the standard deviation of the unexplained regional wage gaps from the Blinder-Oaxaca decomposition increased from 0.13-0.17 in 1993-1998, to 0.24 in 2002 and declined to 0.19 in 2004.<sup>33</sup> Hence, the critical difference is that we observe *global divergence but local convergence*. This suggests that labor markets are becoming regionally more integrated in Vietnam but only locally. If this process of local integration continues, labor markets will also becoming globally integrated over time.

### 4. Labor Market Integration under Market Imperfections

One may argue that the above analysis misses an important point, namely that the major labor market issue in Vietnam is not (regional) wage integration but job creation, especially in rural areas. The current labor force is increasing by approximately 2.6% or 1.3 million each year with most of the increase occurring in rural areas (Le et al. 2003). Employment creation in rural labor markets however has been too weak to absorb this growing labor force and with the increasing rural-urban income gap there is increasing rural-urban migration pressure (Dang et al. 2003). Unless rural labor markets are further developed, this will result in continuing large migration flows from rural to urban areas as well as persistent rural poverty. Hence, from a policy point of view, the focus should be on 'employment' rather than 'wages'. However, one should note that 'employment' and 'wages' are two sides of the same coin. If there are too few jobs then wages tend to be low and vice-versa. This is certainly the case in Vietnam where wages are still very low and labor supply is abundant. However labor markets in Vietnam are segmented (ADB 2005) and market wages may not reflect the actual value of time of people. It is also well known that rural areas in Vietnam suffer from severe underemployment and people are often unable to find jobs at the prevailing market wage. Under these circumstances, a more relevant indicator of labor markets may not be the market wage but the shadow wage. The shadow wage indicates the marginal value of time at the household (or individual) level and will differ from the actual market wage under market imperfections (Sadoulet and De Janvry 1995). In a situation of labor surplus, the shadow price of labor will be below the actual market wage and people are 'trapped' in relatively unproductive activities.

In this section we will therefore estimate *regional shadow wages* for Vietnam to provide a better measure of labor market integration that takes into account the existing (labor and non-labor) market imperfections. We will find that shadow wages are significantly lower than market wages, confirming a lack of integration between wage- and self-employment and the existence of surplus labor in rural areas. However, shadow wages as a ration of market wages have increased between 2002 and 2004 for

<sup>&</sup>lt;sup>32</sup> While the standard deviation of the (raw) wage gap shows divergence: it increases from 0.19-0.23 in 1993-1998 to 0.32 in 2002 and decreases to 0.26 in 2004.

<sup>&</sup>lt;sup>33</sup> Excluding the regions North West and Central Highlands.

the whole country, suggesting that the markets for wage and self-employment are increasingly becoming integrated in Vietnam. In the remainder of this section we first discuss the measurement of regional shadow wages. Next we analyze regional labor market integration on the basis of our measured shadow wages.

### Measurement of regional shadow wages

Most of the labor force in Vietnam consists of farm self-employment and most of Vietnam's labor surplus is found in agriculture. The marginal productivity of farm labor is therefore a good measure of the real value of time in Vietnam. The marginal productivity of farm labor can be calculated from an agricultural production function as its first-order derivative with respect to labor. A number of studies have therefore estimated agricultural production functions to derive household-specific shadow wages (Jacoby 1993, Skoufias 1994). In this paper we follow the same approach, but unlike previous studies, our main interest is not household-specific shadow wages but the regional variation across these wages.

The estimated agricultural production function has the Translog specification<sup>34</sup>:

 $\ln Y = \alpha + \beta \ln X + Z\gamma + \varepsilon$ 

(7)

where *Y* is the output of crop production, *X* is a vector of inputs *including square and interaction terms*, *Z* is a vector of farm, household and community characteristics, and  $\varepsilon$  is an error term. The model has been estimated for 2002 and 2004, and for the panel sample from the two surveys at the household-level.<sup>35</sup> Crop output was measured as the total monetary value of all crops produced in the past 12 months. We did not include outputs from livestock production in our output measure because crop and livestock production have presumably quite different technologies.

As inputs we include measures for land, labor, seeds, fertilizers, insecticides, small tools and other inputs. Land is defined as total land area in squared meters that the household actually cultivated in the last 12 months. Land is calculated by the multiplication of cultivated area of each crop and the number of croppings in the last 12 months. Cultivated land areas of foodstuff and annual industrial crops have been reported as squared meters. Cultivation areas of perennial industrial and fruit crops have been reported as either squared meters or the number of trees. In the case of the latter, the number of trees was converted into an estimated number of squared meters cultivated.<sup>36</sup>

Labor consists of family labor and hired labor. In VHLSS, family labor is measured by the number of working hours that was spent on agricultural activities over the last twelve months. A distinction was also made between male and female family labor and we include both measures in our regressions to control for possible different productivity of male and female labor. Hired labor is measured by the amount of money that the household paid for and this amount has been converted into annual hourly labor input based on the estimated hourly agricultural wage at the province level.

1. The value of each crop is calculated.

<sup>&</sup>lt;sup>34</sup> We tested whether a Cobb Douglas specification was also appropriate but found that some of the interaction terms were strongly significant. However, we also note that the estimated shadow wages are quite similar across the Cobb Douglas and Translog specifications.

<sup>&</sup>lt;sup>35</sup> It was not possible to estimate the model at the plot-level because inputs and outputs have only been measured at the household-level.

<sup>&</sup>lt;sup>36</sup> The conversion was done by the following procedure:

<sup>2.</sup> Yields of each crop (each tree) at household level, district level, provincial level, regional level and country level are computed based on the households reporting the cultivation area in squared meters.

<sup>3.</sup> For those households who reported the cultivation area in the number of trees, we calculate the number of squared meters by taking the values of each crop divided by its yield at district level. If it's still missing (meaning that no households in the district reported the area in squared meters), we will use the yield at next level for which it is available (provincial, regional or country level)

It should be noted that the amount of labor has been measured for all agricultural activities, and that it is not possible to separate labor for cultivation from husbandry activities. In order to correct for this bias, we also include in the regressions the percentage of income from crop production over the total income of crop production and husbandry activities (see Appendix 1 - I for how to correct for measurement error).

Besides land and labor which are considered to be the most important factors for crop production, we also control for other inputs, namely i) seeds, ii) fertilizers, iii) insecticide, iv) small tools, and v) other inputs. All these inputs are measured by their expenditure values over the last twelve months.

Because we estimate a Translog production model we also include square and interaction terms for the most important inputs, namely land, male and female family labor, and hired labor. We chose not to include square and interaction terms for the other inputs to avoid overfitting.

Apart from the input variables (*X*) the production model also includes controls for farm, household and community characteristics (*Z*). As farm characteristics we include a measure for farm quality proxied by the proportion of land that is irrigated. However this information is only available in the VHLSS 2004, and therefore not included in the regressions for 2002. Household characteristics are captured by characteristics of the household head (age in years, gender, and highest official educational degree), as well as age and education composition of the household. The age composition variables measures the proportion of household members whose ages belong to a particular age range, namely 0-15; 15-25; 25-35; 35-45; 45-55; 55-65; and above 65 years. The education levels. Finally, dummies for each province are included to control for differences in climate (such as rainfall) as well as province-differences in input prices.<sup>37</sup>

Because we regard land and labor as essential inputs for the agricultural production process, we have estimated the production function only for households for which both land and at least one of the labor inputs are positive.<sup>38</sup> However, the other inputs are regarded as non-essential and households with zero values for, say, fertilizer, are included in the sample. This immediately creates a problem for our Cobb-Douglass (as well as for the more general Translog) specification, which assumes that all inputs are essential and households with non-essential inputs must be excluded from the sample (because of the logarithmic specification). To avoid this problem, often a "sufficiently small" value to the non-essential inputs is added. Soloaga (2000) points out that this correction is arbitrary and forces the production to include input quantities that are not actually observed. Battese (1997), however, suggests that the problem of zero-inputs can be solved by the use of a dummy such that efficient estimators are obtained but no bias is introduced. Following Battese (1997), assume that the production model with one output and two inputs takes the form of

$$\ln Y_{i} = \alpha_{0} + \beta_{1} \ln X_{1i} + \beta_{2} \ln X_{2i} + \varepsilon_{i} \qquad i = 1, 2, ..., n_{1}$$
(8)

and

$$\ln Y_i = \alpha_1 + \beta_1 \ln X_{1i} + \varepsilon_i \qquad i = n_1 + 1, ..., n_1 + n_2 = n$$
(9)

where  $n_1$  is the number of observations for which  $X_2 > 0$ ,  $n_2$  is the number of observations for which  $X_2 = 0$ , and  $\varepsilon_i$  are the error terms. When pooling the data, the production model can be rewritten as

$$\ln Y_{i} = \alpha_{0} + (\alpha_{1} - \alpha_{0})D_{2i} + \beta_{1}\ln X_{1i} + \beta_{2}\ln X_{2i}^{*} + \varepsilon_{i} \qquad i = 1, 2, ..., n$$
(10)

where  $D_{2i} = 1$  if  $X_{2i} = 0$ ,  $D_{2i} = 0$  if  $X_{2i} > 0$ , and  $X_{2i}^* = \max(D_{2i}, X_{2i})$ .

In the case there are multiple inputs with zero values, one might want to include a dummy

<sup>&</sup>lt;sup>37</sup> The nonessential inputs (seeds, fertilizers, insecticide, small tools, and other inputs) are measured in monetary values and the province dummies will correct for possible price heterogeneity across provinces in input prices.

<sup>&</sup>lt;sup>38</sup> For 2002 1536 observations (7.6%) were excluded and for 2004 101 observations (1.6%) were excluded.

variable for each possible combination of zero inputs in the production model. With n input variables this would give 2<sup>n</sup>-1 dummy variables and possible over-fitting of the regression. We will therefore initially include all these dummy variables but only retain the dummy variables that are significant.

The model has been estimated for 2002 and 2004 both separately as a cross-section as well as a panel for 2002-2004. The descriptive statistics of the model variables are reported in appendix B. In principle 4008 households are available in the 2002 and 2004 panel dataset but only 2471 households report crop production on cultivated land. And after dropping all households with no reported labor input only 2424 households remain.

Table 3 reports the estimates of the agricultural production function for 2002. We present a number of specifications. Specification (1) includes land and family labor inputs, as well as the correction factor for the family labor input variable (proportion of crop income) and province dummies. Both the land and family labor inputs have a positive and significant impact on crop output.<sup>39</sup> Female family labor is somewhat more productive than male family labor and the difference is statistically significant (p-value 0.008). <sup>40</sup> Specification (2) also includes the variable for hired labor and the coefficient for hired labor suggests that hired labor has a positive and significant impact on output. In specification (3) we also include the variables for the other inputs, and this affects the estimated coefficients for labor, suggesting that labor is typically combined with other inputs.

The remaining specifications (4)-(7) add controls for community characteristics (distance to equator) and household characteristics (characteristics of head of household, and household age and education composition) but these have little impact on the estimates for the input variables. Table 4 reports the estimates for 2004 and we note that the results are remarkably similar. Because for 2004 we have also a measure for the proportion of land that is irrigated we have one additional specification including this variable (specification (8)) but the variable has the wrong sign suggesting that it is not a good proxy for land quality. However, most importantly for our analysis, the coefficients for the inputs are barely affected by the inclusion of this variable.

<sup>&</sup>lt;sup>39</sup> Evaluated at the sample mean.

<sup>&</sup>lt;sup>40</sup> Evaluated at the average of the sample mean of male and female family labor input.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sown area in log	0.026	0.136**	0.153***	0.153***	0.157***	0.145***	0.151***
	[0.061]	[0.066]	[0.057]	[0.057]	[0.057]	[0.056]	[0.057]
Sown area in log squared	0.047***	0.039***	0.013***	0.013***	0.012***	0.013***	0.013***
	[0.004]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Proportion of crop income in log	0 214***	0 192***	0 137***	0 137***	0 143***	0.147***	0.154***
	[0.017]	[0.014]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Working hours of male members in log	0.097***	0 112***	-0.084	-0.084	-0.08	-0.078	-0.081
tronking hours of male memorie in log	[0.019]	[0.017]	[0.052]	[0.052]	[0.053]	[0.053]	[0.053]
Working hours of male members in log	[0.0.0]	[0:017]	[0:00_]	[0:00_]	[0.000]	[]	[]
squared	0.007***	0.006***	0.012***	0.012***	0.011***	0.011***	0.011***
	[0.001]	[0.001]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Working hours of female members in log	0.340***	0.277***	0.095**	0.095**	0.088**	0.098**	0.098**
	[0.063]	[0.061]	[0.044]	[0.044]	[0.044]	[0.044]	[0.045]
Working hours of female members in log							
squared	-0.016***	-0.010**	-0.002	-0.002	-0.001	-0.002	-0.002
	[0.004]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Approximate hours of hired labor in log		0.206***	0.096***	0.096***	0.095***	0.095***	0.097***
		[0.037]	[0.031]	[0.031]	[0.031]	[0.032]	[0.032]
Approximate hours of hired labor in log						0.000****	0.000***
squared		0.041***	0.022***	0.022***	0.022***	0.022***	0.023***
		[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Log of area*log of male labor	-0.015***	-0.015***	-0.005***	-0.005***	-0.004***	-0.004***	-0.004***
	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Log of area*log of female labor	-0.008***	-0.008***	-0.003*	-0.003*	-0.003*	-0.003**	-0.003**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Log of area^log of hired labor		-0.042^^^	-0.025^^^	-0.025^^^	-0.025^^^	-0.026***	-0.026***
		[0.005]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Expenses on seeds in log			0.084^^^	0.084^^^	0.084^^^	0.083***	0.081***
-			[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Expenses on fertilizer in log			0.245***	0.245***	0.242***	0.242***	0.238***
			[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Expenses on insecticide in log			0.059***	0.059***	0.059***	0.059***	0.058***
			[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Expenses on small tools in log			0.052***	0.052***	0.052***	0.050***	0.051***
			[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Expenses on other inputs in log			0.143***	0.143***	0.141***	0.140***	0.138***
			[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Community characteristics	No	No	No	Yes	Yes	Yes	Yes
Characteristics of household head	No	No	No	No	Yes	Yes	Yes
Age composition	No	No	No	No	No	Yes	Yes
Education composition	No	No	No	No	No	No	Yes
Province fixed-effects	Yes						
Observations	18559	18559	18559	18559	18559	18559	18559
R-squared	0.75	0.78	0.89	0.89	0.89	0.89	0.89

## Table 3: Agricultural Production Function Estimates (Dependent Variable: Log of Crop Income), 2002

*Note*: Robust standard errors in brackets, \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1% *Source*: Own calculations based on 2002 VHLSS.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sown area in log	-0.037	0.019	0.049	0.049	0.051	0.052*	0.056*	0.062**
	[0.034]	[0.036]	[0.032]	[0.032]	[0.032]	[0.032]	[0.032]	[0.031]
Sown area in log squared	0.051***	0.043***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Proportion of crop income in log	0.175***	0.163***	0.116***	0.116***	0.120***	0.121***	0.125***	0.127***
	[0.019]	[0.018]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Working hours of male members in log	0.117***	0.137***	0.054***	0.054***	0.051***	0.052***	0.051***	0.053***
	[0.025]	[0.025]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]
Working hours of male members in log squared	0.008***	0.007***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Working hours of female members in log	0.325***	0.257***	0.169***	0.169***	0.169***	0.176***	0.169***	0.170***
	[0.072]	[0.070]	[0.062]	[0.062]	[0.062]	[0.062]	[0.062]	[0.062]
Working hours of female members in log squared	-0.014**	-0.009	-0.009*	-0.009*	-0.008*	-0.009*	-0.008*	-0.008*
	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Approximate hours of hired labor in log		0.082**	-0.018	-0.018	-0.02	-0.025	-0.03	-0.027
Approximate hours of hired labor in log		[0.035] 0.038***	[0.036] 0.028***	[0.036] 0.029***	[0.036] 0.028***	[0.036] 0.029***	[0.036] 0.029***	[0.036] 0.030***
Squared		[0.002]	[0.003]	[0.004]	[0.003]	[0.003]	[0.004]	[0.003]
l og of area*log of male labor	-0.017***	-0.018***	-0.010***	-0.010***	-0.010***	-0.010***	-0.010***	-0.010***
	[0.003]	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
l og of area*log of female labor	-0.010***	-0.007**	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002
	[0 003]	[0 003]	[0 002]	[0 002]	[0 002]	[0 002]	[0 002]	[0 002]
l og of area*log of hired labor	[0.000]	-0.025***	-0 019***	-0.019***	-0.019***	-0 019***	-0.019***	-0 020***
		[0 005]	[0 003]	[0 003]	[0.003]	[0 003]	[0.003]	[0 003]
Expanses on seeds in log		[0.000]	0 113***	0 113***	0 11/***	0 113***	0 112***	0 113***
Expenses on seeds in log			[0 010]	[0 010]	[0.010]	[0 010]	[0 010]	[0 010]
Expenses on fertilizer in log			0.206***	0.206***	0.203***	0.203***	0.200***	0.202***
Expenses of fernizer in log			0.230 [0.012]	0.230 [0.012]	0.230 [0.012]	0.230 [0.012]	0.230 [0.012]	[0.232 [0.012]
Exponses on insecticide in log			0.012	0.012***	0.012	0.01/	0.01/	0.046***
			0.045 [0.007]	0.043 [0.007]	0.043 [0.007]	0.044	0.044	0.040
Expanses on small tools in log			0.060***	0.060***	0.060***	0.050***	0.050***	0.057***
Expenses on small tools in log			0.000		0.000	0.039	0.039	0.037
Evenness on other inputs in log			0 102***	0 102***	0 100***	0 100***	0 101***	0 10/***
Expenses on other inputs in log			10 0001	10.123	10.122	10 0001	10.0001	10 0091
Proportion of land irrigated			[0.008]	[0.008]	[0.006]	[0.008]	[0.006]	-0.065***
Community characteristics	No	No	No	Yee	Yee	Yee	Yee	[0.017] Yee
Characteristics of household head	No	No	No	No	Vec	Vec	Vee	Voc
Age composition	No	No	No	No	No	Yes	Yes	Yes
Education composition	No	No	No	No	No	No	Yes	Yes
Province fixed-effects	Yee	Yee	Yee	Yee	Yee	Yee	Yee	Yee
Observations	6149	6149	6149	6149	6149	6149	6149	6149
B-squared	0.79	0.82	0.91	0.91	0.91	0.91	0.91	0.91

# Table 4: Agricultural Production Function Estimates(Dependent Variable: Log of Crop Income), 2004

Note: Robust standard errors in brackets, \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1% Source: Own calculations based on 2004 VHLSS

Finally in Table 5 we present the results for the panel data set for 2002-2004. The Hausman-test is highly significant, suggesting that household random effects should be rejected in favor of household fixed effects. This suggests that the cross-section estimates in tables 3 and 4 may suffer from omitted variable bias because of unobserved household heterogeneity that is correlated with the regressors. The coefficients in the panel regressions are indeed different from those in the cross section regressions, and therefore we will also analyze whether this changes the results for the shadow wages.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sown area in log	-0 117***	-0.091**	-0 116***	-0 121***	-0 120***	-0 122***	-0 114***
	[0.041]	[0.043]	[0.038]	[0.038]	[0.039]	[0.039]	[0.038]
Sown area in log squared	0.048***	0.046***	0.028***	0.029***	0.028***	0.028***	0.028***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Proportion of crop income in loa	0.225***	0.220***	0.172***	0.176***	0.177***	0.175***	0.170***
	[0.022]	[0.021]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]
Working hours of male members						• •	
in log	0.150***	0.180***	0.112***	0.110***	0.106***	0.105***	0.112***
	[0.025]	[0.026]	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]
Working hours of male members							
in log squared	0.003*	0.003	0.002	0.002	0.002	0.002	0.001
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Working hours of female members							
in log	0.273***	0.253***	-0.016	-0.012	-0.012	-0.01	-0.014
	[0.094]	[0.092]	[0.022]	[0.023]	[0.023]	[0.023]	[0.023]
Working hours of female members							
in log squared	-0.016**	-0.014**	0.004**	0.003**	0.003**	0.003**	0.003**
	[0.007]	[0.007]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Approximate hours of hired labor		0.047		0.004			
in log		0.047	0.023	0.024	0.022	0.022	0.022
Annual installation of history labor		[0.035]	[0.029]	[0.029]	[0.029]	[0.029]	[0.029]
Approximate nours of nired labor		0 000***	0 01 1 ***	0 01 1 ***	0 01 1 ***	0 01 1 ***	0 01 1 ***
in log squared		0.020	0.011	0.011	0.011	0.011	0.011
Log of prostlog of male labor	0.010***	0.001***	0.012***	[U.UU2] 0.012***	0.002]	0.012***	0.002]
	0.019	1 20.0- 10 0021	0.013 [0.002]	-0.013 [0.002]	-0.012 [0.002]	-0.012 [0.002]	-0.013 [0.002]
l og of area*log of female labor	0.0003	0.000		_0.002j		0.002	0.002
Log of area log of remain labor	1200.01	10 0031	10 0021	-0.001 [0.002]	10 0021	10 0021	[0.000
l og of area*log of hired labor	[0.000]	-0 014***	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**
		[0 004]	[0 003]	[0 003]	[0 003]	[0 003]	[0 003]
Expanses on seeds in log		[0.004]	0.083***	0.000	0.083***	0.083***	0.081***
Expenses on seeds in log			0.000	0.000 [0.010]	0.000	0.000	0.001
Exponent on fortilizer in log			0.010	0.011***	0.01/***	0.01/***	0.011***
Expenses on lennizer in log			0.210	0.211	0.214	0.214	0.211
European an incenticiale in law							
Expenses on insecticide in log			0.058	0.056	0.05/***	0.050	0.050
<b>F</b>			[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Expenses on small tools in log			0.043***	0.043***	0.042***	0.042***	0.043***
			[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Expenses on other inputs in log			0.095***	0.093***	0.091***	0.090***	0.095***
			[0.009]	[0.009]	[0.009]	[0.009]	[0.009]

### Table 5: Agricultural Production Function Estimates(Dependent Variable: Log of Crop Income), Panel Estimates 2002-2004

48

#### Regional Labor Market Integration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Community characteristics	No	No	No	Yes	Yes	Yes	Yes
Characteristics of household head	No	No	No	No	Yes	Yes	Yes
Age composition	No	No	No	No	No	Yes	Yes
Education composition	No	No	No	No	No	No	Yes
Household fixed-effects	Yes						
Province fixed-effects	No						
Observations	4848	4848	4848	4848	4848	4848	4848
Number of panelid	2424	2424	2424	2424	2424	2424	2424
R-squared	0.54	0.56	0.71	0.71	0.71	0.71	0.71
P_value of Hausman test	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Robust standard errors in brackets, \* significant at 10%; \*\*\* significant at 5%; \*\*\* significant at 1% Source: Own calculations based on 2002 and 2004 VHLSSs

### Regional differences in shadow wages

Based on the comparable final regression results of tables 3-4 (columns 7) we calculated the shadow wage for male and female labor as the expected marginal product of male respectively female labor in each household. We also calculated the average expected market wage for all farmers in the sample based on the Mincer regression results in table 1. Both the predicted shadow wages and expected market wages were corrected for regional price differences. We excluded Hanoi and HCMC from the analysis because of lack of observations (farmers) in these areas.

Table 6 reports the shadow and market wages for males and females in 2002 and 2004 where all wages have been normalized by the mean market wage by gender in each year.<sup>41</sup> The following can be noted from the table. First, shadow wages are significantly below market wages, suggesting that the agricultural self-employment sector suffers from significant surplus labor and that the wage and self-employment sectors remain poorly integrated in Vietnam. Second, shadow wages (as a ratio of market wages) have increased between 2002 and 2004 for the whole country. For males the ratio has increased from 0.09 in 2002 to 0.10 in 2004, while for females the ratio increased from 0.08 in 2002 to 0.09 in 2004. This suggests that the markets for wage and self-employment are increasingly becoming integrated in Vietnam.

Regions	VHLSS 2002				VHLSS 2004			
	М	Male		Female		Male		Female
	Market	Shadow	Market	Shadow	Market	Shadow	Market	Shadow
Red River Delta	0.94	0.08	0.96	0.07	0.97	0.11	1.02	0.09
North East	0.93	0.06	0.96	0.05	0.94	0.07	0.98	0.04
North West	0.84	0.05	0.82	0.04	0.86	0.05	0.79	0.03
North Central Coast	0.95	0.06	0.99	0.05	0.94	0.08	1.00	0.05
South Central Coast	1.04	0.06	1.03	0.06	1.06	0.09	1.02	0.08
Central Highlands	0.95	0.09	0.93	0.09	1.03	0.13	0.99	0.12
South East	1.13	0.11	1.12	0.12	1.16	0.13	1.14	0.15
Mekong River Delta	1.11	0.16	1.10	0.22	1.09	0.17	1.04	0.28
Mean hourly wage	1.00	0.09	1.00	0.08	1.00	0.10	1.00	0.09

### Table 6: Market and Shadow Hourly Wages in Agriculture across Regions in Vietnam,2002-2004

Note: Wages are normalized by mean hourly market wage by gender in each year Source: Own calculations based on 2002 and 2004 VHLSSs

<sup>41</sup> The standard errors of the estimated regional shadow wages are small (typically around 0.01 and 0.04 at most).

Third, shadow wages (as a ratio of market wages) are the lowest in the North East and the North West, and also have been increasing the least in these two regions between 2002 and 2004. This suggests that the integration between the wage and self-employment sectors is the weakest in the North East and the North West. On the other hand the integration is the strongest in the South East and Mekong River Delta and has been further strengthened during 2002-2004.

The above analysis is done for the cross-section estimates of the Translog production functions (Tables 3 and 4). However, the results are very similar if we use the panel estimates instead (Table 5) and the conclusions remain valid.

### 5. Conclusions

In this paper we have provided an in-depth analysis of regional integration of labor markets in Vietnam. Based on the evidence from the household surveys in the period 1993-2004 the paper shows that regional labor markets have been diverging rather than converging during this period. This is also true if one corrects for (changes in) regional differences in human capital, industrial structure and ownership.

However, the paper also suggests that labor market integration is occurring in Vietnam but more locally rather than globally. First, the regional wage gap with neighboring regions is converging, after controlling for human capital, industrial structure and ownership. Second, shadow wages as a ratio of market wages are increasing in most, but not all, regions. This evidence suggests that regional integration is occurring in Vietnam, but primarily at a local scale, and therefore insufficient to accomplish regional integration in Vietnam as a whole. However, with increasing local integration we can also expect global integration in Vietnam, and possibly the observed convergence of regional wage levels during 2002-2004 already reflects this. Evidence from future household surveys will be able to tell whether the observed local convergence will translate into global convergence.

In terms of policy, the analysis in this paper points to a number of issues that could be addressed. First, labor market integration in the North East and North West is extremely weak and should be strengthened in order to improve living standards, reduce poverty, and migration pressures in these regions. Second, shadow wages remain extremely low relatively to market wages in most regions, suggesting that further development of rural labor markets should remain a high priority in Vietnam. Third, and finally, global integration is following local integration, and future policies may want to focus on strengthening the local rather than national linkages between regions.

#### Acknowledgements

We would like to thank Henrik Hansen for helpful discussions.

### References

- ADB. 2005. "Labor market segmentation and poverty." A report for the ADB MMW4P Project" (IWP, CIEM and ILSSA). Draft report.
- Bernard, A. and S. Durlauf. 1996. "Interpreting tests of the convergence hypothesis." *Journal of Econometrics*, 71: 161-73.
- Battese, G. E. 1997. "A Note on the Estimation of Cobb-Douglas production functions when some Explanatory variables have Zero values", *Journal of Agricultural Economics*, 48(2): 250-252.
- Dang, N. A., C. Tacoli, and X. T. Hoang. 2003. "Migration in Vietnam. A review of information on current trends and patterns, and their policy implications". Research paper, RMMRU and DFID, Dahka.
- Gallup, J. L. 2002. "Wage labour markets and inequality in Vietnam in the 1990s", In P. Glewwe, N.

Agrawal and D. Dollar, *Economic Growth, Poverty, and Household Welfare in Vietnam*. Regional and Sectoral Studies, Washington DC: World Bank

- Jacoby, H. 1993. "Shadow wages and Peasant Family Labor Supply: An Econometric Application to the Peruvian Sierra", *Review of Economic Studies*, 60: 903-21.
- Klump, R. and N. Anh. 2004. "Patterns of provincial growth in Vietnam, 1995-2000: Empirical analysis and policy recommendations", Working paper.
- Le, X. B., K. D. Nguyen Thi, and H. H. Tran. 2003. "Some issues relating to the development of the labour market in Vietnam", Central Institute for Economic Management, Hanoi.
- Liu, A. 2001. "Markets, Inequality and Poverty in Vietnam", Asian Economic Journal, 15(2): 217-35.
- Nguyen, Cong Gian, H. Q. Doan, L. H. Nguyen Thi, and R. Oostendorp. 2006. "Trade Liberalization, The Gender Wage Gap And Returns To Education In Vietnam", Vietnam Economic Research Network
- Quah, D. 1996. "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics", *The Economic Journal*, 106: 1045-55.
- Rassekh, F. 1998. "The convergence hypothesis: History, theory, and evidence", Open
- Economics Review, 9: 85-105.
- Rassekh, F. and H. Thompson. 1998. "Micro convergence and macro convergence: Factor price equalization and per capita income", *Pacific Economic Review*, 3(1): 3-11.
- Sadoulet, E. and A. De Janvry. 1995. *Quantitative Development Policy Analysis*. The Johns Hopkins University Press, Baltimore and London.
- Sala-i-Martin, X. 1996. "The classical approach to convergence analysis", *Economic Journal*, 106: 1019-36.
- Skoufias, E. 1994. "Using Shadow Wages to Estimate Labor Supply of Agricultural Households", American Journal of Agricultural Economics, 76(2): 215-227
- Soloaga, I. 2000. "The Treatment of Non-Essential Inputs in a Cobb-Douglas Technology. An application to Mexican rural household level data", *Policy Research Working Paper 2499*, Washington DC: World Bank.
- Solow, R. 1956. "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, 70(1): 65-94.
- World Bank. 2004. Vietnam Development report 2005: Governance. Hanoi, Vietnam.
- World Bank. 2005. Vietnam Development report 2006: Business. Hanoi, Vietnam.

### **Appendix 1. Mathematical Formulations**

### I. Formulation for Correcting Measurement Error

Suppose the model without the measurement error can be described by the equation below:

[1]  $\log(Y_c) = \alpha + \beta \log(L_c) + Z\gamma + \varepsilon$ 

where  $Y_c$  is crop production,  $L_c$  is crop production labor, Z is a vector of other control variables and  $\varepsilon$  is an error term. Let  $\overline{L}$  denote the total crop production and husbandry labor. The equation [1] is equivalent to the following one:

[2]  $\log(Y_c) = \alpha + \beta \log(\overline{L}) + Z\gamma + (\varepsilon + \beta \log(L_c / \overline{L}))$ 

which can be rewritten as

[3]  $\log(Y_c) = \alpha + \beta \log(L) + Z\gamma + \eta$ , with  $\eta = \varepsilon + \beta \log(L_c/\overline{L})$ 

There are three potential solutions to correct the measurement error problem:

1. include additional variables to control for  $\log(L_C / \overline{L})$ , such as the share of total production from different crops and from different types of livestock.

2. use instruments for  $\log(\overline{L})$ , and possibly Z, that are uncorrelated with  $\log(L_C/\overline{L})$ .

3. do panel analysis assuming that  $\log(L_C / \overline{L})$  is constant over time. We adopt solution (1) (in the cross section analysis) as well as solution (3) (in the panel analysis). We do not apply solution (2) because there are no obvious instruments.

### II. Formulation of Expected Shadow Wage

The expected shadow wage is given by  $E[\frac{\partial Y}{\partial L_i}] = E[\beta \frac{Y}{L_i}] = E[\beta \frac{\exp(\alpha + \beta \ln X + \gamma Z + \varepsilon)}{L_i}] = E[\beta \frac{\exp(\alpha + \beta \ln X + \gamma Z}{L_i}]E[e^{\varepsilon}],$ 

which has been estimated by  $\hat{\beta}[\frac{\exp(\hat{\alpha}+\hat{\beta}\ln X+iZ)}{L_i}][\frac{1}{N_i}\sum e^{\hat{\varepsilon}}]$ , with *i* standing for male or female and  $N_i$  the total number of households with positive (female or male) labor input.

### Appendix 2. Descriptive Statistics of the Variables in the Mincer Models

Table 1: Summary Statistics of Variables Based on VLSS 1993 Data

Variable	Obs	Mean	Std. Dev	Min	Max
log hourly wage	1863	0.45	0.77	-1.64	2.99
Female	1863	0.43	0.49	0.00	1.00
Years of schooling	1863	8.88	3.71	0.00	18.00
Female x years of schooling	1863	3.86	5.08	0.00	18.00
Experience	1863	16.49	10.43	0.00	55.00
Experience squared (10-3)	1863	0.38	0.46	0.00	3.03
Mining	1863	0.01	0.11	0.00	1.00
Manufacturing	1863	0.32	0.47	0.00	1.00
Electricity, construction	1863	0.09	0.29	0.00	1.00
Commerce	1863	0.06	0.23	0.00	1.00
Transportation, communication	1863	0.05	0.21	0.00	1.00
Finance, other services	1863	0.30	0.46	0.00	1.00
Government	1863	0.25	0.43	0.00	1.00
SOE	1863	0.18	0.38	0.00	1.00
FDI	1863	0.01	0.08	0.00	1.00
Red R Delta	1863	0.12	0.33	0.00	1.00
NorthEast	1863	0.09	0.29	0.00	1.00
NorthWest	1863	0.01	0.08	0.00	1.00
N Central Coast	1863	0.07	0.26	0.00	1.00
S Central Coast	1863	0.12	0.32	0.00	1.00
Central Highland	1863	0.01	0.09	0.00	1.00
Southeast	1863	0.09	0.29	0.00	1.00
Mekong Delta	1863	0.23	0.42	0.00	1.00
Hanoi	1863	0.08	0.26	0.00	1.00
HCMC	1863	0.18	0.39	0.00	1.00

Source: Own calculations based on 1992/93 VLSS.

Variable	Obs	Mean	Std. Dev	Min	Max
log hourly wage	2244	1.40	0.62	-0.29	3.48
Female	2244	0.39	0.49	0.00	1.00
Years of schooling	2244	7.70	3.91	0.00	22.00
Female x years of schooling	2244	2.86	4.33	0.00	19.00
Experience	2244	17.14	10.63	0.00	58.00
Experience squared (10-3)	2244	0.41	0.48	0.00	3.36
Mining	2244	0.02	0.14	0.00	1.00
Manufacturing	2244	0.39	0.49	0.00	1.00
Electricity, construction	2244	0.16	0.37	0.00	1.00
Commerce	2244	0.10	0.30	0.00	1.00
Transportation, communication	2244	0.08	0.27	0.00	1.00
Finance, other services	2244	0.01	0.12	0.00	1.00
Government	2244	0.04	0.19	0.00	1.00
SOE	2244	0.22	0.41	0.00	1.00
FDI	2244	0.09	0.29	0.00	1.00
Red R Delta	2244	0.10	0.30	0.00	1.00
NorthEast	2244	0.06	0.23	0.00	1.00
NorthWest	2244	0.00	0.07	0.00	1.00
N Central Coast	2244	0.08	0.27	0.00	1.00
S Central Coast	2244	0.12	0.33	0.00	1.00
Central Highland	2244	0.01	0.08	0.00	1.00
Southeast	2244	0.17	0.38	0.00	1.00
Mekong Delta	2244	0.19	0.39	0.00	1.00
Hanoi	2244	0.05	0.22	0.00	1.00
HCMC	2244	0.22	0.42	0.00	1.00

Table 2: Summary Statistics of Variables Based on VLSS 1998 Data

Source: Own calculations based on 1997/98 VLSS.

Variable	Obs	Mean	Std. Dev	Min	Max
Log hourly wage	19194	1.22	0.59	-0.49	3.08
Female	19194	0.38	0.49	0.00	1.00
Years of schooling	19194	8.24	4.94	1.00	22.00
Female x years of schooling	19194	3.17	5.15	0.00	22.00
Experience	19194	18.58	10.92	0.00	56.00
Experience squared (10-3)	19194	0.46	0.49	0.00	3.14
Mining	19194	0.02	0.16	0.00	1.00
Manufacturing	19194	0.23	0.42	0.00	1.00
Electricity, construction	19194	0.16	0.36	0.00	1.00
Commerce	19194	0.07	0.25	0.00	1.00
Transportation, communication	19194	0.05	0.22	0.00	1.00
Finance, other services	19194	0.25	0.43	0.00	1.00
Government	19194	0.21	0.41	0.00	1.00
SOE	19194	0.14	0.34	0.00	1.00
FDI	19194	0.03	0.16	0.00	1.00
Red R Delta	19194	0.16	0.37	0.00	1.00
NorthEast	19194	0.10	0.30	0.00	1.00
NorthWest	19194	0.02	0.13	0.00	1.00
N Central Coast	19194	0.08	0.26	0.00	1.00
S Central Coast	19194	0.11	0.31	0.00	1.00
Central Highland	19194	0.04	0.20	0.00	1.00
Southeast	19194	0.14	0.35	0.00	1.00
Mekong Delta	19194	0.26	0.44	0.00	1.00
Hanoi	19194	0.05	0.21	0.00	1.00
HCMC	19194	0.05	0.22	0.00	1.00

Table 3: Summary	Statistics of	f Variables	Based on	VHLSS	2002	Data

Source: Own calculations based on 2002 VHLSS.
			Std.		
Variable	Obs	Mean	Dev	Min	Max
log hourly wage	5778	1.47	0.58	-0.20	3.04
Female	5778	0.38	0.48	0.00	1.00
Years of schooling	5778	9.16	4.18	0.00	22.00
Female x years of schooling	5778	3.57	5.31	0.00	22.00
Experience	5778	18.22	11.17	0.00	58.00
Experience squared (10-3)	5778	0.46	0.48	0.00	3.36
Mining	5778	0.02	0.14	0.00	1.00
Manufacturing	5778	0.25	0.43	0.00	1.00
Electricity, construction	5778	0.18	0.38	0.00	1.00
Commerce	5778	0.08	0.27	0.00	1.00
Transportation, communication	5778	0.06	0.23	0.00	1.00
Finance, other services	5778	0.29	0.45	0.00	1.00
Government	5778	0.25	0.43	0.00	1.00
SOE	5778	0.13	0.33	0.00	1.00
FDI	5778	0.05	0.21	0.00	1.00
Red R Delta	5778	0.20	0.40	0.00	1.00
NorthEast	5778	0.11	0.32	0.00	1.00
NorthWest	5778	0.02	0.15	0.00	1.00
N Central Coast	5778	0.08	0.28	0.00	1.00
S Central Coast	5778	0.11	0.31	0.00	1.00
Central Highland	5778	0.04	0.19	0.00	1.00
Southeast	5778	0.12	0.33	0.00	1.00
Mekong Delta	5778	0.19	0.39	0.00	1.00
Hanoi	5778	0.05	0.23	0.00	1.00
HCMC	5778	0.07	0.25	0.00	1.00

Table 4: Summary Statistics of Variables Based on VHLSS 2004 Data

Source: Own calculations based on 2004 VHLSS.

# Appendix 3. Descriptive Statistics of the Variables in the Agricultural Production Models

	Unit	Obs	Mean	s.d	Min	Max
Output						
Values of crop production	000 VND	18559	8678.04	10952.81	10.32	199085.84
Inputs						
Land sown area	M2	18559	9607.82	12964.36	4.00	590560.00
# working hours of males		12662	1785.39	1258.10	6.00	15888.00
# working hours of females		15351	1856.71	1265.30	4.00	12432.00
# working hours of hired labor		8905	705.12	1832.49	4.84	152969.67
Expenses on seeds	000 VND	17124	410.96	828.94	2.04	45454.86
Expenses on fertilizers	000 VND	18023	1136.61	2017.25	4.03	183563.61
Expenses on insecticide	000 VND	16436	333.25	774.23	1.04	18529.37
Expenses on small tools	000 VND	14958	83.23	154.96	1.99	3377.75
Expenses on other inputs	000 VND	17476	1040.82	1764.44	2.87	71385.85
Household characteristics						
Highest education degree of head		18559	1.32	1.26	0.00	6.00
Age of head in years	Years	18559	47.70	14.18	17.00	102.00
Percentage of member age in 0-15	%	18559	30.32	21.98	0.00	85.71
Percentage of member age in 15-25	%	18559	18.09	20.30	0.00	100.00
Percentage of member age in 25-35	%	18559	14.17	19.59	0.00	100.00
Percentage of member age in 35-45	%	18559	13.69	18.49	0.00	100.00
Percentage of member age in 45-55	%	18559	9.12	17.77	0.00	100.00
Percentage of member age in 55-65	%	18559	6.10	16.06	0.00	100.00
Percentage of member age over 65	%	18559	8.51	20.52	0.00	100.00
Percentage of non-educated people	%	18559	40.06	30.62	0.00	100.00
Percentage of primary-educated people	%	18559	27.63	24.69	0.00	100.00
Percentage of secondary-educated people	%	18559	23.19	25.28	0.00	100.00
Percentage of high-school-educated people	%	18559	6.30	13.50	0.00	100.00
Percentage of vocationally educated people	%	18559	0.60	4.36	0.00	100.00
Percentage of professionally educated people	%	18559	1.39	7.09	0.00	100.00
Percentage of university-educated people	%	18559	0.82	5.19	0.00	100.00

Table 1: Summary Statistics of Variables Based on VHLSS 2002 Data

*Note:* Mean weighted by sampling weights. Obs column shows the number of positive observation in the dataset. *Source:* Own calculations based on 2002 VHLSS.

	Unit	Obs	Mean	s.d	Min	Max
Output						
Values of crop production	000 VND	6249	9795.38	13630.38	17.81	367625.63
Inputs						
Land sown area	M2	6249	9158.49	12030.88	0.10	200000.00
# working hours of males		5153	1310.44	1105.81	10.00	12564.00
# working hours of females		5640	1494.88	1151.26	1.00	10336.00
# working hours of hired labor		3028	984.64	3137.47	7.96	126164.66
Expenses on seeds	000 VND	5674	481.70	1948.86	0.96	95983.93
Expenses on fertilizers	000 VND	6076	1570.02	2618.48	0.99	54538.10
Expenses on insecticide	000 VND	5522	466.55	1140.28	3.02	24312.27
Expenses on small tools	000 VND	5441	87.18	176.13	0.97	7417.53
Expenses on other inputs	000 VND	5753	1194.28	2125.03	1.91	45347.23
Household characteristics						
Highest education degree of head		6249	1.51	1.44	0.00	6.00
Age of head in years	Years	6249	48.93	13.77	15.00	98.00
Percentage of member age in 0-15	%	6249	27.75	21.87	0.00	80.00
Percentage of member age in 15-25	%	6249	18.41	20.44	0.00	100.00
Percentage of member age in 25-35	%	6249	13.00	18.64	0.00	100.00
Percentage of member age in 35-45	%	6249	14.38	19.04	0.00	100.00
Percentage of member age in 45-55	%	6249	11.16	19.52	0.00	100.00
Percentage of member age in 55-65	%	6249	6.73	17.19	0.00	100.00
Percentage of member age over 65	%	6249	8.57	20.16	0.00	100.00
Percentage of non-educated people	%	6249	35.31	30.20	0.00	100.00
Percentage of primary-educated people	%	6249	27.47	24.71	0.00	100.00
Percentage of secondary-educated people	%	6249	23.85	25.82	0.00	100.00
Percentage of high-school-educated people	%	6249	7.18	14.35	0.00	100.00
Percentage of vocationally educated people	%	6249	2.39	9.01	0.00	100.00
Percentage of professionally educated people	%	6249	2.45	9.29	0.00	100.00
Percentage of university-educated people	%	6249	1.35	7.11	0.00	100.00

Table 2: Summary Statistics of Variables Based on VHLSS 2004 Data

Note: Mean weighted by sampling weights. Obs column shows the number of positive observations in the dataset. Source: Own calculations based on 2004 VHLSS.

# 3

# Determinants and Impacts of Migration in Vietnam

### Nguyen Thi Thu Phương, Tran Ngo Thi Minh Tam, Nguyen Thi Nguyet and Remco Oostendorp

This chapter uses the recent Vietnam Household Living Standard Survey 2004 to analyze the determinants and impacts of migration in Vietnam. Most of the previous studies on the determinants and impacts of migration have focused on destination rather than origin areas of migration. This limits our understanding of the determinants of migration and also does not provide evidence on important impacts of migration such as on household inequality in origin areas.

In terms of determinants of migration, the study shows that migration is a highly selective process and strongly affected by household and commune characteristics, although differently across type of migration and across urban and rural areas. We do find evidence for the existence of a 'migration hump' for economic long-term migration, with an inverted U-shape in the probability of migration with respect to per capita expenditures. The presence of non-farm employment opportunities does reduce short-term migration but not long-term out-migration for economic reasons.

In terms of impacts the study analyzes the impact of migration on household expenditures and household inequality. Migration is found to have a strong positive impact on household expenditures but increases the Gini coefficient of per capita household expenditures from 0.38 to 0.42 in origin areas compared to the no-migration case.

#### 1. Introduction

Migration issues have been given great attention in Vietnam. Several studies have been carried out to analyze migration patterns along with other aspects of socio-economic development in Vietnam (Guest 1998, Djamba *et al.* 1999, Goldstein *et al.* 2001, GSO and UNFPA 2005, and Dang 2005). Other studies examine the determinants of migration (Trinh 1998, Dang and Le 2001, Dang *et al.* 1997, 2006, and Nguyen, T. L. 2001), its consequences (Do and Nguyen 1998, Le, V.T. 1998, and Nguyen, D.V. 2001) and other related issues such as fertility (White *et al.* 2000), social capital (Dang 1998), gender gap (Dang and Le 2001), livelihoods (Dang *et al.* 2004), and environment (WWF 1999).

Studies on migration in Vietnam have been based on nationwide statistics (census and administrative data), large-scale survey data and small-scale survey data (including case studies). The census data studies are based on the Population and Housing Census in 1989 (Dang *et al.* 1997) and

1999 (GSO and UNDP 2001). They provide characteristics of permanent migrants<sup>42</sup> but exclude shortterm, unofficial, and seasonal migrants. Also because census data are collected only once every 10 years, they often cannot provide up-to-date information. Apart from census data some studies have also used administrative data on migration from the Department for Resettlement and New Economic Zones (Do, V.H. 1998). According to these nationwide statistics, 1.6 people million moved from rural to other rural areas, 1.13 million people moved between urban centers, and 1.18 million people moved from rural settlements to urban centers, while over 400,000 moved in the opposite direction, from urban centers to rural areas between 1994 and 1999 (Dang *et al.* 2003).

Migration studies using large-scale surveys have been based on the Viet Nam Demographic and Health Survey in 1997 (VNDHS 1997), the Viet Nam Migration and Health Survey in 1997 (VNMHS 1997), the Vietnam Migration Survey in 2004 (VMS 2004), and the Vietnam Living Standard Survey 1998 (VLSS 1998). The earlier large-scale surveys, such as the Vietnam Living Standard Survey in 1993 and the Viet Nam Inter-Censal Demographic Survey (VNICDS) in 1994 and VNMHS in 1996 did not ask about migration. The VNDHS 1997 was the first large-scale survey which allowed researchers to define migrants versus non-migrants but it was only conducted for married women.<sup>43</sup> The VNMHS 1997 and VMS 2004 included migrants and non-migrants aged 15 and above and many studies have been based on these data sources (White *et al.* 2000, Dang and Le, 2001, Goldstein *et al.*, 2001, Nguyen, D.V., 2001, and Nguyen and White 2002, Dang and Nguyen 2006). These surveys provide detailed pictures of both in- and out-migration flows in terms of socio-economic characteristics of the (non-) migrants, reasons for moving and not-moving, history of occupational and residential mobility, the process of migration, employment, income, urban integration, and health-related characteristics.

However, both the VNMHS 1997 and VMS 2004 are not fully representative for Vietnam as they are limited to a small number of provinces that are known to have high rates of in- and/or outmigration. The VNMHS 1997 surveyed 2502 individuals in two rural origin areas (Ha Nam and Thai Binh) and in four destination areas (Hanoi, Da Nang, Dak Lak, and Binh Duong). The VMS 2004 contained a much larger sample of 10.000 individuals but was only conducted in five destination areas, namely a number of large cities and economic zones.

The most representative large-scale survey that has been used in migration studies so far is the Vietnam Living Standard Survey 1998 (VLSS 1998). The Vietnam Living Standard Surveys are considered to be the most representative surveys for the whole population of the country containing very comprehensive information on household and commune background characteristics. However the Vietnam Living Standard Surveys of 1993 and 2002 have no information on migration status of the respondents and therefore current migration studies have been based on the VLSS 1998 only. Also it has been noted that even the Vietnam Living Standard Surveys cannot be regarded as fully representative, because of under-sampling of newly-formed households, and not properly incorporating migrants without residency permits in Hanoi and Ho Chi Minh City (UNDP and GSO 2001, Gallup 2002, ADB 2005, and Dang *et al.* 2005). Nevertheless, because of the availability of valuable information on individual and household characteristics like employment, income, expenditure, or assets, the VLSS 1998 has been an important source of information for migration studies in Vietnam (GSO 2000, Le, X.B. 2001, and Nguyen, T. 2002).

Apart from the studies based on nationwide statistics and large-scale surveys, several migration studies have been based on small-scale surveys and case studies (Truong *et al.* 1996, Doan *et al.* 1998, Do and Nguyen 1998, Guest 1998, Dang *et al.* 2005, Nguyen 2005). These studies provide important and often in-depth information on migration characteristics and its impacts on origin and destination areas. Moreover, a number of studies have analyzed migration in relation to other aspects such as

<sup>&</sup>lt;sup>42</sup> Defined as residing in a different province at the time of survey compared to 5 years ago.

<sup>&</sup>lt;sup>43</sup> Also the survey was limited to six provinces where large migration flows were observed.

transport planning (WB and MOT 2006), labor market segmentation (ADB 2005), and wage income (Hanoi Trade Union and ActionAid 2004, Oostendorp 2004)

In this study we will use the most recent Vietnam Household and Living Standard Survey of 2004 (VHLSS 2004) to study the determinants and impacts of migration. The VHLSS 2004 is an interesting source to study migration for a number of reasons. First, the VHLSS 2004 is representative for the whole of Vietnam, unlike the recent Vietnam Migration Survey 2004, and has not been used to study migration before. The representativeness of the survey is a big advantage as it allows us to study the determinants of migration by comparing the characteristics of migrants and non-migrants in the sending rather than destination areas. As will be discussed later in the chapter, earlier studies on the determinants of migration in Vietnam have typically compared migrants with non-migrants in destination areas and this is potentially quite misleading given that the non-migrants in sending and receiving areas may be quite different.

Second, because the VHLSS 2004 forms a panel data set with the VHLSS 2002, it is possible to study how migration status is related to household characteristics *before* the change in migration status. Therefore our analysis will suffer less from potential endogeneity problems unlike the previous studies in Vietnam that were based on cross-sectional analyses. Third, as noted before, the VHLSS 2004 is extremely comprehensive in terms of individual, household and community-level information, and therefore our analysis will be able to control for a great number of factors in our analysis of the determinants and impacts of migration.

However, the VHLSS 2004 has also a number of disadvantages when studying migration issues. First, the sampling frame is based on official residence records. Therefore selected households are mainly those households which were already granted official permanent residence registration in the province (99.3%). One or two single migrants moving into the area for a temporary purpose or even a long-term residence are therefore typically not included. The household's official permanent residence status stamped by the local authority has many requirements which are very difficult for migrants to fulfill, especially in big cities. For example, a household moving into Hanoi for at least 5 years could be considered to get the KT1 – an official permanent residence status – only by the time of 2004. Second, because the VHLSS 2004 was not designed specifically for the purpose of studying migration, the information on migration is more limited than what can be found in the migration surveys VNMHS 1997 and VMS 2004. Nevertheless, and in spite of these limitations, we believe that the VHLSS 2004 remains an extremely useful source of data to improve our understanding of the migration process in Vietnam.

The remainder of this chapter is structured as follows. In section 2 we first discuss the different approaches to measuring migration flows in the literature. Next we discuss the migration patterns in Vietnam, both in terms of existing evidence from previous studies as well as new evidence from the VHLSS 2004.

In section 3 we analyze the determinants of migration in Vietnam. After a discussion of different theoretical models of migration we turn to the empirical analysis of actual determinants of short-term and long-term migration in Vietnam. Special attention will be given to the possible existence of a 'migration hump' in the sense that the probability of migration has an inverted U-shape with respect to the income status of the household.

In section 4 we turn to the analysis of the impacts of migration. Most studies of migration have looked at the impacts of migration on the migrants and/or the receiving areas. Very few studies have looked at the impacts of migration on the sending areas. In this study we will analyze the impact of migration of household members on their household, in terms of remittances, household expenditures, and inter-household inequality. Earlier studies have pointed out the important role of remittances for the economic welfare of remaining household members, but there is no previous study that has provided a systematic study of the impact of migration on household expenditures and inter-household inequality. Finally, section 5 summarizes the results.

#### 2. Migration Patterns in Vietnam

#### 2.1. The definition and measurement of migration

It is important to note that there is no uniformity in definitions and measures of migrations used in the literature. The main reasons for this are twofold. First, migration is difficult to define because it involves both a time and spatial dimension which need to be defined carefully in turn. Second, migration studies often use different definitions because they rely on different data sources.

Let us first turn to the difficulties in defining migration itself. Bilsborrow (1996) has provided a very careful discussion of these issues and here we summarize his main points. Migration involves both the time and spatial dimension, because it is, most generally, viewed as a movement in space during a certain period of time. However, both 'movement in space' and 'period of time' need to be further specified because not all movements are automatically regarded as migration. One conventional definition of migration considers any movement a migration if (1) it involves the crossing of a political or administrative boundary, and (2) a change in "usual residence". This definition was particularly developed because it allows for the measurement of migration flows from typical population census data.

It is not hard to see the limitations of this conventional definition. First, although trivial moves such as moving next door clearly should not be counted as migration, political and administrative boundaries are arbitrary, changeable and not necessarily stable over time. Therefore any move may be counted as 'migration' within one institutional context but not in another context.<sup>44</sup> Second, the requirement that migration involves a change in 'usual residence' excludes less permanent types of migration that are increasingly recognized to be important as well. Many people are involved in 'temporary' or 'seasonal' migration without changing their usual place of residence. Also some people are involved in 'circular' migration which has been defined in some studies as 'sleeping a series of consecutive nights away from home, and is repeated on a more-or-less regular basis in consecutive months, but without changing one's perceived place of usual residence" (Bilsborrow 1996, p.5).

Apart from the difficulty of defining migration, the second reason why migration studies have been using different definitions for migration is simply that they have used different sources of data. For instance Dang *et al.* (1997) defined a migrant as someone of five years or older who 'moved into a different district/province' in the previous five years as this was the information available in the population census data. Other data sources often lend themselves to different measures of mobility. One study using the VNMHS 1997 identified a 'rural to urban migrant' as someone who 'left the rural communes for the urban places between 1987 and 1997 who were at least 15 years old when moved' but excluding 'those migrants who reported marriage as the reason for move' (Dang 2001). Or another study using the VMS 2004 defined a migrant as 'a person aged 15-59 years who had moved to their current district/quarter from another district/quarter during the five years prior to the survey, and who had resided at their current place of residence one month or more' except for a person 'who had moved from one quarter to another within Hanoi or HCM City' (GSO 2005).

Given the diversity in migration definitions, as well as the constraints imposed by any data source, it is important to define explicitly our definition(s) of migration before turning to our empirical analysis of the determinants and impacts of migration in Vietnam. Before doing this, however, we first discuss the existing evidence on migration patterns in Vietnam.

<sup>&</sup>lt;sup>44</sup> For instance because of the sheer size of provinces in China, any long-distance within-province movement in China would not be counted as 'migration' while a similar long-distance move in a country with smaller provinces will probably be counted as 'migration'.

#### 2.2. Previous evidence on migration patterns in Vietnam

Although there are differences in the definition of migration, studies on internal migration in Vietnam consistently point towards increasing rates of migration. The next table presents the net migration flows for the period 2002-2004 for the five most important provinces of destination and origin of migration in Vietnam. The information on migration in the table is based on the Census of 1999 and the annual GSO survey of the demographic changes. Migration has been defined as the movement of people out of the province for any living purpose. The five most important destination areas were Ho Chi Minh, Hanoi, Binh Duong, Quang Ninh and Da Nang, while the most important origin areas were Thanh Hoa, Nam Dinh, Thai Binh, Ha Tay and Quang Nam. However among the five most important destination areas, the migration inflow was highly concentrated and mostly directed at Ho Chi Minh, Hanoi, and Binh Duong. The net migration outflow was more equally spread among the five most important origin areas. This suggests that Ho Chi Minh, Hanoi and Binh Duong are the dominant magnets for internal migration for a large range of provinces. Table 1 also shows that migrants tend to move from provinces with low GDP per capita levels, low Human Development Index (HDI) and high underemployment rates to provinces with high GDP per capita levels, high HDI and low underemployment rates.

Table 1 presents migration flows from respectively to the most important origin and destination areas, but not the total aggregate migration flow in Vietnam. The aggregate flow can only be measured with nationwide statistics or representative large-scale surveys. With respect to the former, census data for 1999 reveal that 6.5 percent of the 69 million persons over five years of age in 1999 lived in a different province in 1994. The corresponding percentage from the 1989 census for the period 1984-89 was 2 percent, showing that the inter-province migration flow increased significantly since the eighties. If one adds intra-province movement (between rural communes or urban wards), then the percentage of people over five years of age who lived in a different place in 1999 compared to five years before was 6.5%. (Dang *et al.* 2003). This amounts to a total of 4.35 million people who moved between 1994-99, of which 1.6 million people moved from rural to other rural areas, 1.13 million people moved between urban centers, and 1.18 million people moved from rural settlements to urban centers, while over 400,000 moved in the opposite direction, from urban centers to rural areas between 1994 and 1999 (Dang *et al.* 2003).

With respect to representative large-scale surveys, the Vietnam Living Standards Survey of 1998 shows that 6.89% of people of 16 years and older moved within the past 5 years. The percentage was higher for people in urban areas compared to rural areas (7.92% versus 6.45%), reflecting a predominance of rural-urban migration flows (Nguyen 2002). Cumulatively, about one-half of urban residents were born in the countryside, against 90% of the rural population (Le *et al.* 2001). So far no studies have used the most recent Vietnam Household Living Standards Survey 2004 to measure aggregate migration flows in Vietnam and this will be done in the next section.

Previous migration studies have also shown that migration is selective in Vietnam. Migrants are often quoted as those who are relatively young, less likely to be married, and disproportionately female.<sup>45</sup> However, little difference in educational levels completed has been found between migrants and non-migrants in destination areas (GSO and UNFPA 2005).<sup>46</sup> This observed selectivity holds for both temporary and permanent migration, although the dominance of young females in migration streams is mostly observed in the long-term migration flow from rural to urban and industrial zones. As for short-term migration, males are more likely to migrate as females tend to be responsible for household work (VASS 2006).

<sup>&</sup>lt;sup>45</sup> Comparison of the 1989 and the 1999 censuses, suggests that the emergence of the female migrant is a recent phenomenon (Dang *et al.*, 2003, p.6)

<sup>&</sup>lt;sup>46</sup> In the next section we will see that evidence from the VHLSS 2004 shows that migrants tend to be better educated than non-migrants in the area of *origin*.

Province	Net migration (persons) 2002- 2004	Migration rank 2002- 2004	Migration rate (%) <sup>-</sup>	GDP per capita** ('000 VND/head) 2002	Rank of GDP per capita 2002	Rank of HDI 2002	Under- employment rate (%) 2002	Rank of under- employment 2002
5 most desti	nations				_	-		
Ho Chi Minh	210,237	1	3.84	11620.70	2	3	6.48	61
Hanoi	148,063	2	5.05	8410.18	3	2	22.63	50
Binh Duong	40,761	3	5.18	6642.85	4	6	17.55	57
Quang Ninh	5,248	4	0.50	4897.46	10	10	17.64	56
Da Nang	3,941	5	0.54	5915.89	6	4	15.77	58
5 most origir	ıs							
Thanh Hoa	-37,848	64	-1.07	2579.20	41	35	79.62	5
Nam Dinh	-27,482	63	-1.42	2653.42	38	17	73.06	12
Thai Binh	-23,350	62	-1.28	2809.06	32	12	79.89	4
На Тау	-19,723	61	-0.80	2771.49	33	24	31.03	43
Quang Nam	-17,455	60	-1.23	2524.89	42	25	71.53	17

Table 1: Net Migration Flows from Respectively to the Most Important Origin
and Destination Areas, 2002-2004

Source: GSO (2006).

Note: \* Percentage of population in 2002; \*\* At constant price.

Although Table 1 suggests that economic reasons are clearly linked to migration flows, migrants move for economic *and* non-economic reasons. Based on the VLSS 1998, it has been found that among migrants who moved between 1993 and 1998 economic reasons are not the strongest ones leading to either rural-to-rural or rural-to-urban migration (22%). The main reasons are related to war (10.6%), family relations (59%) and others (8.5%) (Le *et al.* 2001). In the next section we will use the VHLSS 2004 to analyze whether the same pattern holds for the most recent migration flows.

#### Evidence on migration from the VHLSS 2004

In this study we will use the VHLSS 2004 which allows for a number of different measures of migration. First, we can define a *long-term migrant since* 2002 as someone of age 15 or above who was a household member in 2002, but who is no longer a household member in 2004 (but still alive).<sup>47</sup> In the VHLSS 2004 a household member is defined as someone who shares lodging, income, and expenditures for at least 6 months in the past 12 months. Hence, someone can become a long-term migrant if he/she either moves away, splits from the household, or stays in the household less than 6 months in the past 12 months.

Second, the VHLSS 2004 allows us to define a *short-term out-migrant in 2004* as someone of age 15 or above who is a member of the household but who has been absent for at least one month.<sup>48</sup>

<sup>&</sup>lt;sup>47</sup> This implies that the migrant was of age 13 years or older in 2002.

<sup>&</sup>lt;sup>48</sup> But for not more than six months because otherwise this person would no longer be counted as a household member and become a long-term out-migrant instead.

Third, we can also define an *in-migrant since* 2002 as someone of age 15 or above who is a household member in 2004 but who was not a household member in 2002.<sup>49</sup>

Fourth, we also note that some studies have used the presence of (domestic and/or international) *remittances* as a proxy for out-migration. Rodriguez (1998), in his study of the impact of international remittances on income and income distribution, uses this proxy because he has no information on the actual migrants in a household. The advantage of this measure is that it also includes migration that may have occurred before the survey period (unlike the usual measures such as those above), but its disadvantage is that it is only a proxy for remitting migrants and misses households with non-remitting migrants. The VMS 2004 suggests that only approximately one-half of the (domestic) Vietnamese migrants remitted money back home during the past 12 months, and therefore this proxy severely underestimates the actual rate of migration. Also the information on remittances in the VHLSS does not only include receipts from former household members but also from other relatives and friends, and hence the receipts of remittances is not always linked to migration. We therefore we limit ourselves in this analysis to the above measures of long-term migrant since 2002, short-term migrant in 2004, and in-migrant.

Before discussing the estimated number and the characteristics of the migrants based on the above definitions, we need to discuss two additional issues. First, it should be noted that all of these definitions only relate to *individual* rather than household migration. The primary sampling unit of the VHLSS 2004 is the household and there is no record of what happened to the VHLSS households that have moved away.<sup>50</sup> This is an important limitation, as we also know from the recent VMS 2004 that approximately 37.7% of the migrants have moved with family (GSO 2005, Table 3.10). Therefore our analysis looks only at the determinants and impacts of *individual* migration in Vietnam.

Second, the above definitions of migration do not include an explicit geographical boundary. This implies that within and between-province movements are both counted as migration.<sup>51</sup> In many studies migration is defined relative to a geographical boundary, such as province borders. Unfortunately, the VHLSS 2004 does not provide information on the location to which a household member has moved. This may be problematic in so far as the determinants and impacts of migration may vary with distance.

Therefore the question arises whether the lack of information on destination area is a serious problem for our analysis. First we note that unlike the VHLSS 2004, the VLSS 98 not only provides information on whether a household member has left the household ('long-term migration'), but also whether this member has moved to another province (or even country). It is found that this is the case for 46% of the household members who moved out between 1993 and 1998. If we are willing to assume that this pattern is not dramatically altered between 2002 and 2004, then we can state that our measure of migration has a strong (but imperfect) correlation with between-province movement.

Second, the probability that a move out of the household also involved a between-province move depends on the characteristics of the household member and the reason for movement. For instance, using the VLSS 98, we find that males are more likely to move to another province than females (50 versus 43%). Also the young (less than 30 years, 48%) and the old household members (more than 50 years, 51%) are more likely to move to another province than the middle-aged (between 30 and 50 years, 38%). The reason for movement also matters - namely 82% of the move involved a between-province move if the reason was schooling, 70% if employment, 66% if other reasons, 45% if follow family, and 36% if marriage or

<sup>&</sup>lt;sup>49</sup> The VHLSS 2004 also includes a question on how long a household member has been living in the current province/city, but only 278 out of 39,696 respondents did reply to this question, and none of these respondents were new household members since 2002. Hence, this information is of no value for measuring (in-)migration.

<sup>&</sup>lt;sup>50</sup> This is unfortunate because in principle it should have been possible to record the reason why households that were interviewed in 2002 were no longer interviewed in 2004. This was done in the VLSS 1998 for the 1993 households, but not in the VHLSS 2004 for the 2002 households.

<sup>&</sup>lt;sup>51</sup> And in some cases trivial moves within the same commune or ward will be included as well.

household split. Therefore, in order to control for the possibility that the determinants and/or impacts of migration may vary across within-province migration and between-province migration, it is useful to test whether the determinants and impacts vary significantly across different types of migrants. If the biggest impact is found for, say, young migrants that move for reasons of employment, then this impact is most likely linked to migration across province borders. If on the other hand, the impact is largest for females between 30 and 40 years old who move because of marriage or household split, then this impact is most likely linked to within-province movement.

Table 2 reports the estimated number of migrants and households with migrants. On average 2.5% of the individuals of age 15 or above was absent for at least one but at most six months in 2004 (short-term migration). In total 10.7% of the individuals of age 15 or above left the household between 2002 and 2004 (long-term migration). And 4.7% of the individuals of age 15 or above moved into an existing household (in-migration). If we look at the percentage of households with migrants, we see that 7.3% of the households had at least one short-term migrant, 26.1% had at least one long-term migrant, and 12.6% has at least one in-migrant.

It was already noted that the VHLSS 2004 does not allow us to distinguish between within- and between-province moves. However, based on the evidence from the VLSS 1998, we can assume that moves for reasons of schooling, employment and other are most likely between-province moves, while moves for family, marriage or household split reasons are most likely within-province moves. Table 2 also reports the percentage of long-term out-migrants by reason for movement (no information on reason for movement is given for short-term and in-migration). Most long-term migrants move out of the household because of household split (3.1%) or marriage (3.0%) and presumably many of these moves are within-province.<sup>52</sup> However, also many long-term migrants move out because of work (2.8%), and these are presumably mostly between-province moves.<sup>53</sup>

	Short-term	Long-term	
	out-migrant	out-migrant	In-migrant
Individuals (%)*			
Migrant	2.5	10.7	4.7
By reason			
Household split		3.1	
Marriage		3.0	
Work		2.8	
Go with family		0.5	
Study		0.3	
Other		1.0	
Households (%)			
Household with migrant	7.3	26.1	12.6
With			
1 migrant	6.0	17.1	9.5
2 migrants	1.0	6.7	2.4
3 migrants	0.2	1.7	0.5
>3 migrants	0.0	0.6	0.2

#### Table 2: Number of Migrants and Households with Migrants with Reasons for Migration (%)

*Note*: Numbers are weighted with sampling weights. Percent of individuals of age 15 or above.

<sup>&</sup>lt;sup>52</sup> But presumably also many of these moves are between-province as we found that for 1998 that 36% of the moves because of marriage or household split was between provinces.

<sup>&</sup>lt;sup>53</sup> For 1998 we found that 70% of the work-related moves were between provinces.

The next table reports the socio-economic characteristics of migrant and non-migrants. For longterm migrants we distinguish between migration because of work ('economic' reasons) and because of other reasons ('non-economic' reasons).<sup>54</sup> The table shows the following. First, economic long-term migrants tend to be relatively young compared to the other migrants and non-migrants. Second, while most of the short-term and economic long-term migrants are male (61.7 respectively 59.2%), most of the non-economic long-term migrants are female (59.8%). Third, economic long-term migrants are far less likely to be married (12.8%) compared to the other migrants and non-migrants. Fourth, migrants tend to have more education than non-migrants, especially short-term and economic long-term migrants. This apparently contradicts the earlier finding that there is little difference in educational levels completed between migrants and non-migrants (GSO and UNFPA 2005). However, this earlier finding was based on a comparison between migrants and non-migrants in destination areas. Here we compare migrants and non-migrants in the areas of origin, and we find that there is positive selection according to education. The breakdown by level of education completed shows that migrants are much more likely to have obtained an upper-secondary school or even a tertiary diploma. Fifth, economic long-term migrants are most likely to originate from rural areas, confirming the earlier findings of a predominance of rural migration. However for short-term migration we do find little difference between urban and rural areas. Finally, economic long-term migrants originate from households with per capita expenditures of 4,815,000 VND, while short-term migrants live in households with per capita expenditures of 5,418,000 VND. This suggests that long-term migrants are more likely to originate in poor households than short-term migrants, although other factors may also play a role.<sup>55</sup> We therefore now turn to an analysis of the determinants of migration.

Information by the time of 2004		Long	Long-term <sup>a,b</sup>		
	Short-		Non-	migrant	
	terma	Economic	Economic		
Age (years)	31	23	28	31	
Male (%)	61.7	59.2	40.2	49.6	
Married (%)	43.9	12.8	39.0	46.4	
Education (years of education)	8.8	9.3	7.8	6.6	
Has no diploma (%)	12.8	10.9	20.3	34.4	
Has primary education as the highest diploma obtained (%)	27.2	24.6	32.0	28.7	
Has lower-secondary education as the highest diploma					
obtained (%)	32.2	37.1	27.1	23.4	
Has upper-secondary education as the highest diploma					
obtained (%)	24.3	21.3	14.5	10.6	
Has tertiary and higher diploma (%)	3.5	6.1	6.1	2.9	
Urban (%)	22.0	13.5	25.0	23.3	
Household size (persons)	4.2	3.9	4.2	4.4	
Per-capita expenditures ('000 VND)	5418	4815	5796	4922	

Table 3: Socio-Economic Char	acteristics of Mi	igrants and No	on-Migrants
------------------------------	-------------------	----------------	-------------

*Note*: Numbers are weighted with sampling weights. The migrants in the table refer to long- and short-term out-migrants (not in-migrants). <sup>a</sup> Information on urban, household size and per-capita expenditures are for household from which migrant originates. <sup>b</sup> Information of education and marital status, are at the time of 2002.

<sup>&</sup>lt;sup>54</sup> It should be noted that migration for 'economic' reasons may also have non-economic motives and migration for 'non-economic' reasons may also include economic reasons. Hence, if preferred, one may also speak of migration for 'work' and 'non-work' reasons.

<sup>&</sup>lt;sup>55</sup> Including the fact that per-capita expenditures will be affected by migration as well.

#### 3. Determinants of Migration in Vietnam

#### 3.1. Economic models of the determinants of migration

The economic literature on the determinants of migration assumes that individuals or households rationally consider various locations and choose the location that maximizes the expected gains from migration. The expected gains of migration potentially depend on a large number of factors and different models have been developed to account for different factors.

In the classical Todaro migration model (1969), the expected gains are measured by (a) the difference in real incomes between job opportunities across locations, and (b) the probability of getting a job in each location. Todaro applied this model to rural-urban migration flows, and showed that urban unemployment can coexist with rural-urban migration flows as long as the expected wage in urban areas is higher than the rural wage. Harris and Todaro also showed that under certain conditions urban job creation programs may actually cause the level and rate of urban unemployment to rise because of increased rural-urban migration (Harris and Todaro 1970).

Subsequent economic models of migration have extended the Todaro framework to address a number of shortcomings in the model and to explain why migration sometimes fails to occur even when substantial expected earnings differentials exist or, conversely, why migration sometimes occurs even when earnings differentials are absent. The basic model assumes that individuals or households have perfect information about job opportunities and earnings differentials across locations. In reality, migrants often rely on networks to acquire information which creates "tied" or "chain" migration (Stark 1991, Lalonde and Topel 1997, Knight and Song 1997). The model assumes that individuals or households maximize expected income instead of expected utility, ignoring such "non-economic" costs and benefits as family ties, cultural differences, and social status (Thadani and Todaro 1984, Stark and Taylor 1989, Stark 1991, Lalonde and Topel 1997). Also income differentials may be less relevant than considerations of security in situations of civil unrest, armed conflicts, political persecution and natural disasters. Uncertainty may also affect migration, as a risk-averse person may choose not to migrate even if the expected earnings differential is positive (Katz and Stark 1986, Lakshmansamy 1990). With imperfect credit markets, migration may generate cash income and alleviate the credit constraint and therefore migration may occur even if the expected income differential is negative (Stark 1980, Collier and Lal 1984). Similarly, the decision to migrate should be viewed as a joint household rather than an individual decision and therefore may depend not only on individual characteristics and preferences but also on the characteristics and preferences of the other household members (Stark 1991). Apart from differences in real incomes between job opportunities across locations, also the availability of government transfers across locations may affect migration (Enchautegui 1997).

Numerous studies have used the (extended) Todaro framework to estimate the determinants of migration. However, the number of studies estimating the determinants of migration is limited in Vietnam and they have been based on the 1989 census (Dang *et al.* 1997), VNMHS97 (Dang and Le 2001, Nguyen, T. L. 2001), VMS 2004 (Dang *et al.* 2006), and a survey on spontaneous migrants in Hanoi (1997), Ho Chi Minh and Vung Tau (1996,1997) (Trinh 1998). Most of these studies are micro studies but compare the characteristics of migrants and non-migrants in *destination* areas. This is potentially very misleading, as non-migrants in origin and destination areas are presumably quite different, and therefore the proper comparison is between the characteristics of migrants and non-migrants in *origin* areas. The study using the 1989 census does not suffer from this weakness but it looks at the determinants of aggregate (provincial) migration flows without controlling for household and migrant characteristics. In the next section we will therefore use the VHLSS 2004 to estimate the determinants of migration in Vietnam. The VHLSS 2004 offers two important advantages compared to

the data sources used in previous studies on the determinants of migration in Vietnam. First, it is a representative large-scale survey, allowing us to compare migrants and non-migrants in *origin* areas. Second, the VHLSS 2004 is part of a panel with the VHLSS 2002, and we can therefore use the characteristics of the households and migrants *prior* to the migration decision to study the determinants of migration. The previous studies were all based on cross-section data and therefore prone to simultaneity bias.

#### 3.2. Estimating the determinants of migration in Vietnam using the VHLSS 2004

We analyze the determinants of migration at the household level and not at the individual level. Migration is typically a joint household decision, and the decision to migrate depends on household characteristics rather than individual characteristics.<sup>56</sup> Apart from household characteristics (X), we also include commune characteristics (Z) among the determinants as migration depends not only on household characteristics but also on the characteristics of the origin and destination areas.

The probability that a household has a short-term or long-term migrant is modeled by two probit models:<sup>57</sup>

 $Pr[\text{household has short-term migrant in 2004}] = \Phi(X_i\beta^{ST} + Z_i\gamma^{ST})$ 

 $Pr[\text{household has long-term migrant in 2002-04}] = \Phi(X_i\beta^{LT} + Z_i\gamma^{LT})$ 

where *i* is the household,  $\Phi$  is the standard normal distribution, and  $\beta$ ,  $\gamma$  are parameters to be estimated.

Among the household characteristics we include household size (measured as number of household members) as we expect that larger households are more likely to have migrants because of surplus labor. Per capita household expenditure (measured in logarithm) is included as a proxy for the household's living standard and because poorer households have a bigger incentive to migrate. However, we also include a square term for per capita household expenditure to test whether the relationship between migration and economic well-being is non-linear. Some studies have suggested the existence of a 'migration hump', in the sense that the poorest of the households are too poor to migrate while the richest have no incentive (Lucas 2005).58 Households with more agricultural land per capita (in hectare) are less likely to have migrants because of a lower labor surplus. Also these households may be better able to access credit and therefore have a lower incentive to use migration to generate liquidity. We also include age composition shares as we expect that households with many young adults are more likely to have migrants. Educational shares are also included as the expected income gain from migration tends to be larger for higher levels of education and therefore households with better educated members are more likely to have migrants. Also better educated households may have better access to information about destination areas. The gender and marital status of the household head are included because we expect that households with married heads are less likely to have (long-term) migrants, while the expected income gain from migration may depend on gender

<sup>&</sup>lt;sup>56</sup> But the household characteristics should reflect the heterogeneity of the household, such as educational and age composition. For instance, a household with four members of age 30 is expected to exhibit very different migration behavior from a household of equal mean age but with, say, two household members of age 10 and two household members of age 50.

<sup>&</sup>lt;sup>57</sup> For reasons of simplicity, we have estimated two separate probit models rather than one bivariate probit model. In econometric terms, this leads to a possible loss in efficiency (but not consistency) as we have ignored the possible correlation in unobserved factors explaining short-term and long-term migration.

<sup>&</sup>lt;sup>58</sup> There is prima facie evidence for the existence of a migration hump as the percentages of short-term and long-term migrants show an inverted U-shape with respect to the per capita household expenditure quintiles.

because of different labor market opportunities.<sup>59</sup> Finally the share of household members working in wage employment and in the private sector<sup>60</sup> is included. We expect that households already involved in wage employment or the private sector are more likely to consider migration as a possible livelihood strategy unlike purely agricultural households or households involved in state employment.

Apart from the household characteristics we also include commune characteristics as the migration decision depends also on the characteristics of the area of origin. First a number of proxies are included for the general level of development of the areas of origin. The poverty status of the commune (a dummy variable for whether the commune is classified as poor in the province) is included as more people may want to migrate if the commune is poorly developed. A dummy variable indicating whether the commune has enterprises, factory or trading village within 10 km is included as people will be less likely to migrate if they have more non-farm job opportunities. Also a dummy variable indicating whether the commune has a job creation program is included. Although one may be skeptical about the efficacy of these programs, we include this indicator to formally test whether it reduces migration pressures. A dummy variable indicating whether the commune has a good climate for agriculture is included (measured as whether natural disasters/drought/flood are among the three main problems faced by farmers) as we expect that communes with low agricultural productivity to face strong pressures for out-migration. The distance to the nearest road accessible for cars (in km) is included to measure the accessibility of the commune. On the one hand poor access will limit the opportunities for market integration and increase migration pressures, while on the other hand it will increase the cost of (especially short-term) migration. A dummy indicating the presence of electricity is included as communes without electricity will face stronger pressures to migrate. Finally a dummy variable indicating whether malaria, leprosy, goiter or tuberculosis are among the three main illnesses in the commune is included as we expect higher out-migration rates in those communes.

Apart from the above household and commune characteristics, a dummy variable for urban area and regional dummies are included to control for other geographical differences affecting the incentive to migrate, such as climate and local opportunities for migration.

It should be noted that all of the above household and commune characteristics have been constructed from the VHLSS 2002 and are therefore defined prior to the migration status as recorded in the VHLSS 2004. This will significantly reduce the problem of simultaneity that has plagued the earlier cross-section studies on migration in Vietnam. At the same time this implies that we will only use the 4008 households that are included in both the VHLSS 2002 and VHLSS 2004.

Because the commune characteristics are only available for rural communes, we will estimate three versions of the models for short- and long-run migration:

- (1) For both rural and urban with only household characteristics (4008 households).
- (2) For rural areas with only household characteristics (3100 households)
- (3) For rural areas with household and commune characteristics (3100 households)

The comparison of (1) versus (2) will tell us whether the determinants of migration vary across rural and urban areas. The comparison of (2) versus (3) will tell us whether commune characteristics have independent explanatory power apart from the household characteristics.

Table 4 presents the descriptive statistics of the model variables, for both urban and rural areas and for rural areas alone. On average 8% of the households have short-term migrants and 26% have long-term migrants. These figures are quite similar to those reported in Table 2, even if the figures in Table 4 are unweighted and only for the panel households.

<sup>&</sup>lt;sup>59</sup> For instance because of the increasing demand for female labor in export-oriented industries (Wood 1991).

<sup>&</sup>lt;sup>60</sup> Domestic or foreign.

VARIABLES	BO L	TH RURAL JRBAN AR	AND EA	ONLY RURAL AREA			
	Mean	Min	Max	Mean	Min	Max	
Urban	0.22	0.00	1.00				
Short-term out-migration	0.08	0.00	1.00	0.08	0.00	1.00	
Long-term out-migration	0.26	0.00	1.00	0.26	0.00	1.00	
Household size (persons)	4.52	1.00	18.00	4.57	1.00	14.00	
Log of Per capita household expenditure ('000 VND)	8.51	6.99	11.28	8.37	6.99	10.28	
Log of Per capita household expenditure ('000 VND)							
squared	72.75	48.91	127.32	70.29	48.91	105.65	
Per capita household agricultural land (hectare)	0.19	0.00	7.13	0.22	0.00	7.13	
Age of less than 15 (as over household size)	0.27	0.00	0.80	0.28	0.00	0.80	
Age of 15 to 25 (as over household size)	0.20	0.00	1.00	0.20	0.00	1.00	
Age of 25 to35 (as over household size)	0.15	0.00	1.00	0.15	0.00	1.00	
Age of 35 to 45 (as over household size)	0.15	0.00	1.00	0.14	0.00	1.00	
Age of 45 to 55 (as over household size)	0.09	0.00	1.00	0.09	0.00	1.00	
Age of above 55 (as over household size)	0.14	0.00	1.00	0.14	0.00	1.00	
No education (as over household size)	0.15	0.00	1.00	0.16	0.00	1.00	
Primary education (as over household size)	0.33	0.00	1.00	0.36	0.00	1.00	
Lower secondary education (as over household size)	0.34	0.00	1.00	0.35	0.00	1.00	
Upper secondary education (as over household size)	0.18	0.00	1.00	0.13	0.00	1.00	
HH head being a male	0.78	0.00	1.00	0.81	0.00	1.00	
HH head being married	0.84	0.00	1.00	0.84	0.00	1.00	
Household wage employment composition (as over							
household size)	0.21	0.00	1.00	0.19	0.00	1.00	
Household employment in private sector (as over							
household size)	0.53	0.00	1.00	0.57	0.00	1.00	
Being poor commune in the province or not				0.22	0.00	1.00	
Having enterprises, factory, trading village in the circular of							
10km				0.65	0.00	1.00	
Having job-creation program in the commune				0.37	0.00	1.00	
Not good climate for agriculture				0.22	0.00	1.00	
Distance to the nearest car road (km)				0.67	0.00	45.00	
Having electricity in the commune or not				0.93	0.00	1.00	
Having any noteworthy disease in the commune or not				0.45	0.00	1.00	
Red River Delta	0.20	0.00	1.00	0.21	0.00	1.00	
North East	0.14	0.00	1.00	0.14	0.00	1.00	
North West	0.04	0.00	1.00	0.04	0.00	1.00	
North Central Coast	0.11	0.00	1.00	0.12	0.00	1.00	
South Central Coast	0.10	0.00	1.00	0.09	0.00	1.00	
Central Highlands	0.06	0.00	1.00	0.06	0.00	1.00	
Southeastern	0.12	0.00	1.00	0.10	0.00	1.00	
Mekong River Delta	0.22	0.00	1.00	0.23	0.00	1.00	
No of observations		4008			3100		

 Table 4: Descriptive Statistics of the Model Variables

In Table 5 the marginal effects of the Probit regression results are presented. First, we find that there are no significant differences in migration flows between urban and rural areas after we control for the other factors. Second, the Red River Delta is most dynamic in terms of short- and long-term migration (respectively 3.7 and 9.6% higher than the reference region Central Highlands). Also longterm migration from North Central Coast is significantly higher at 11.3%. Third, larger households tend to have more migrants and this is especially the case for long-term migrants. Fourth, we do not find any significant relationship between household income status (proxied by per capita household expenditure) and land holdings and migration. Fifth, age has a large impact on migration behavior, as households with members of age between 15 and 25 years are much more likely to have out-migrants (5.7-7.9% higher probability of having a short-term migrant and 22.6-24.7% higher probability of having a long-term migrant). Sixth, education increases the probability of out-migration significantly. A household where all members have upper secondary education has a 12.3-12.7% higher probability to have short-term migrants, and a 16.5-17.9% higher probability to have long-term migrants compared to a household where all members have no education. Households involved in wage employment and the private sector show significant higher probability of migration. Finally, the final two regressions in Table 5 show the impact of commune characteristics on out-migration in rural areas. We find that communes with enterprises, factories, or trading village within 10 km have significantly less out-migration. This suggests that local economic development of non-farm opportunities is important to reduce migration pressures. At the same time the presence of a jobcreation program does not appear to have been successful in reducing this pressure. Puzzlingly, we also find that communes without a good climate for agriculture face lower rates of long-term outmigration. Maybe after a natural disaster/drought/flood people are less likely to migrate because of reconstruction efforts. In summary, we can conclude that migration is a highly selective process and that migration pressures will remain high as long as the population is young, increasingly educated, and non-farm economic opportunities are lacking.

	ONLY	( HOUSEHOLD	BOTH HOUSEHOLD AND COMMUNE CHARACTERISTICS					
Probability of	BOTH U	JRBAN	ONLY RURAL AREA					
Out-migration	AND RUR	Al Areas						
	Short-term	Long-term	Short-term	Long-term	Short-term	Long-term		
	Out-migration	Out-migration	Out-migration	Out-migration	Out-migration	Out-migration		
Urban	0.0004 [0.03]	-0.019 [0.96]						
Red River Delta	0.037	0.096	0.073	0.136	0.078	0.131		
	[1.97]**	[2.67]***	[2.92]***	[3.01]***	[3.08]***	[2.90]***		
North East	-0.008	0.029	-0.002	0.049	-0.005	0.041		
	[0.49]	[0.83]	[0.09]	[1.10]	[0.24]	[0.92]		
North West	0.034	0.061	0.026	0.046	0.012	0.019		
	[1.33]	[1.28]	[0.87]	[0.84]	[0.39]	[0.34]		
North Central Coast	-0.026 [1.59]	0.113	-0.008 [0.36]	0.133 [2.79]***	-0.007 [0.34]	0.129 [2.69]***		

 Table 5: Probit Regressions of Determinants of Short- and Long-Term

 Out-Migration by Area, 2004 (Marginal Effects)

	ONLY HOUSEHOLD CHARACTERISTICS				BOTH HOUSEHOLD AND COMMUNE CHARACTERISTICS		
Probability of Out-migration	BOTH U AND RUR/	JRBAN AL AREAS		ONLY R	JRAL AREA		
	Short-term	Long-term	Short-term	Long-term	Short-term	Long-term	
	Out-migration	Out-migration	Out-migration	Out-migration	Out-migration	Out-migration	
South Central Coast	-0.004	0.054	0.01	0.1	0.014	0.102	
	[0.23]	[1.41]	[0.39]	[2.01]**	[0.59]	[2.05]**	
Southeastern	-0.042	-0.008	-0.03	-0.003	-0.025	-0.001	
	[2.61]***	[0.21]	[1.44]	[0.06]	[1.21]	[0.02]	
Mekong River Delta	-0.031	0.027	-0.022	0.038	-0.015	0.042	
	[1.97]**	[0.80]	[1.16]	[0.91]	[0.78]	[0.99]	
Household size	0.006	0.091	0.006	0.094	0.006	0.095	
	[2.06]**	[17.34]***	[1.98]**	[15.55]***	[2.10]**	[15.44]***	
Per capita household	0.156	0.43	0.037	0.401	0.075	0.452	
expenditure	[0.98]	[1.63]	[0.15]	[0.90]	[0.29]	[1.00]	
Per capita household	-0.01	-0.024	-0.003	-0.023	-0.006	-0.025	
expenditure squared	[1.10]	[1.62]	[0.24]	[0.87]	[0.36]	[0.95]	
Per capita household	-0.003	-0.026	0.005	-0.022	0.004	-0.028	
agricultural land	[0.27]	[1.19]	[0.41]	[1.03]	[0.31]	[1.30]	
Age of less than 15	0.025	-0.433	0.049	-0.406	0.05	-0.413	
	[0.88]	[8.17]***	[1.38]	[6.39]***	[1.42]	[6.49]***	
Age of 15 to 25	0.057	0.226	0.079	0.247	0.074	0.228	
	[1.97]**	[4.23]***	[2.42]**	[4.37]***	[2.27]**	[4.01]***	
Age of 25 to35	-0.019	0.02	-0.018	0.041	-0.019	0.03	
	[0.63]	[0.35]	[0.51]	[0.70]	[0.55]	[0.50]	
Age of 35 to 45	-0.014	-0.017	-0.026	0.028	-0.027	0.019	
	[0.50]	[0.31]	[0.80]	[0.48]	[0.82]	[0.32]	
Age of 45 to 55	-0.015	0.026	-0.015	0.022	-0.012	0.02	
	[0.52]	[0.51]	[0.52]	[0.41]	[0.42]	[0.37]	
Primary education	0.088	0.062	0.09	0.043	0.09	0.049	
	[3.29]***	[1.38]	[3.14]***	[0.88]	[3.20]***	[1.00]	
Lower secondary	0.11	0.115	0.121	0.06	0.122	0.072	
education	[4.03]***	[2.49]**	[3.99]***	[1.20]	[4.08]***	[1.41]	
Upper secondary	0.127	0.165	0.123	0.167	0.124	0.179	
education	[4.07]***	[3.06]***	[3.24]***	[2.78]***	[3.33]***	[2.95]***	
HH head male	0.007	-0.05	0.0002	-0.045	0.0003	-0.049	
	[0.57]	[2.18]**	[0.01]	[1.47]	[0.02]	[1.61]	
HH head married	-0.023	-0.045	-0.014	-0.046	-0.014	-0.043	
	[1.49]	[1.76]*	[0.81]	[1.40]	[0.84]	[1.33]	
Household wage	0.03	0.066	0.031	0.069	0.036	0.078	
employment	[1.73]*	[2.03]**	[1.52]	[1.85]*	[1.75]*	[2.07]**	
composition							

	ONLY	( HOUSEHOLD	BOTH HOUSEHOLD AND COMMUNE CHARACTERISTICS					
Probability of BOTH URBAN Out-migration AND RURAL AREAS			ONLY RURAL AREA					
	Short-term Out-migration	Long-term Out-migration	Short-term Out-migration	Long-term Out-migration	Short-term Out-migration	Long-term Out-migration		
Household employment in private sector	0.017 [0.92]	0.063 [1.85]*	0.031 [1.35]	0.141 [3.33]***	0.028 [1.26]	0.141 [3.32]***		
Being poor commune in the province or not					-0.012 [1.05]	-0.001 [0.06]		
Having enterprises, factory, trading village in the circular of 10km					-0.021 [2.07]**	-0.044 [2.34]**		
Having job-creation program in the commune					-0.003 [0.33]	0.019 [1.10]		
Not good climate for agriculture					0.001 [0.08]	-0.038 [1.98]**		
Distance to the nearest car way					-0.002 [1.36]	-0.001 [0.49]		
Electricity					-0.033 [1.46]	-0.039 [1.02]		
Noteworthy disease					0.004	0.024 [1.49]		
Observations	4008	4008	3100	3100	3100	3100		

Absolute value of z statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

It may be argued that the determinants of migration are different for different types of migration. For instance, one may distinguish between within- and between-province migration. Or between migration for economic and non-economic reasons. We have already discussed that the VHLSS 2004 does not provide information on where the short- or long-term migrant has moved, and hence we cannot study the determinants of within- versus between-province migration. However, we know the reason for long-term migration, and evidence from the VLSS 1998 suggests that most of the migrants (70%) who move because of work ('economic' reason) tend to move to another province, while most of the migrants (61%) who move because of non-work reasons ('non-economic' reasons) tend to move within the province. We therefore also estimate the determinants of long-term migration for economic and non-economic reasons separately (Table 6). The determinants for economic long-term migration are strikingly different in some respects from those for non-economic long-term migration. First, long-term migration for economic reasons is more likely a rural than urban phenomenon. Second, economic long-term migration is more likely from Red River Delta, North East, North Central Coast and Mekong River Delta. Third, although household size increases the probability of long-term out-migration, the impact is the largest for out-migration for noneconomic reasons. Fourth, we now find a clear inverted U-shape ('migration hump') for long-term

out-migration for economic reasons with respect to economic well-being in rural areas.<sup>61</sup> This suggests that the poorest are less likely to migrate for economic reasons, presumably because of the costs and uncertain benefits involved. Fifth, while also the age group 15-25 years is more likely to migrate for economic reasons, we now also find a significant larger probability of economic long-term out-migration for people of age 45 to 55 years in rural areas. This suggests that economic pressures are also strong on the older people in rural areas and that they choose to migrate for economic reasons after the children have grown older. Finally, with respect to the commune characteristics, we find that communes with enterprises, factories or trading villages within 10 km do experience less non-economic out-migration but not less economic out-migration. This together with the fact that communes with job creation programs do not face less migration pressure, suggests that economic out-migration is mostly driven by the economic opportunities in destination areas rather than origin areas.

	ONLY	Y HOUSEHOLD	STICS	BOTH HOUSEHOLD AND COMMUNE CHARACTERISTICS				
Probability of	BOTH AND RUR	URBAN Al Areas		ONLY RURAL AREA				
out-ingration	Economic Long-term Out-migration	Non- Economic Long-term Out-migration	Economic Non-Economic Long-term Long-term Out-migration Out-migration		Economic Long-term Out-migration	Non-Economic Long-term Out- migration		
Urban	-0.017 [2.07]**	-0.01 [0.53]						
Red River Delta	0.07	0.049	0.057 [1.95]*	0.094 [2.43]**	0.051 [1.82]*	0.092 [2.36]**		
North East	0.042	-0.001	0.045	0.009	0.049	0.0004		
North West	0.024	0.036	-0.012	0.032	-0.005	0.003		
North Central Coast	0.15	0.004	0.153	0.004	0.15	-0.001		
South Central Coast	0.029	0.029	0.032	0.07 [1.64]*	0.029	0.073		
Southeastern	0.042	-0.029 [0.95]	0.04 [1.25]	-0.022 [0.58]	0.038	-0.02 [0.53]		
Mekong River Delta	0.07 [2.79]***	-0.015 [0.52]	0.063 [2.20]**	-0.001 [0.03]	0.06 [2.12]**	0.004 [0.12]		
Household size	0.014 [6.69]***	0.077 [16.66]***	0.017 [6.88]***	0.078 [15.03]***	0.017 [6.90]***	0.078 [14.93]***		

 
 Table 6: Probit Regressions of Determinants of Economic and Non-Economic Long-Term Out-Migration by Area, 2004 (Marginal Effects)

<sup>&</sup>lt;sup>61</sup> The probability of economic long-term out-migration is increasing in per capita expenditure levels below 4,674,000 or 3,878,000 in the regressions without and with commune characteristics respectively, and decreasing afterwards.

	ONLY	ONLY HOUSEHOLD CHARACTERISTICS BOTH HOUSEHOLD COMMUNE CHARACTER					
Probability of Out-migration	BOTH U AND RUR	JRBAN Al Areas	ONLY RURAL AREA				
	Economic Long-term Out-migrationNon- Economic Long-term Out-migrationEconomic Long-term Out-migrationNon-Economic Long-term Out-migration		Economic Long-term Out-migration	Non-Economic Long-term Out- migration			
Per capita household	0.183	0.312	0.507	0.055	0.462	0.13	
expenditure	[1.49]	[1.33]	[2.38]**	[0.14]	[2.17]**	[0.33]	
Per capita household	-0.011	-0.017	-0.03	-0.002	-0.028	-0.006	
expenditure squared	[1.58]	[1.27]	[2.43]**	[0.07]	[2.23]**	[0.24]	
Per capita household	-0.009	-0.018	-0.011	-0.013	-0.01	-0.019	
agricultural land	[0.92]	[0.95]	[1.21]	[0.70]	[1.15]	[0.99]	
Age of less than 15	-0.047	-0.393	-0.042	-0.373	-0.041	-0.378	
	[1.89]*	[8.22]***	[1.38]	[6.72]***	[1.39]	[6.83]***	
Age of 15 to 25	0.068	0.154	0.067	0.163	0.069	0.144	
	[2.81]***	[3.21]***	[2.63]***	[3.27]***	[2.72]***	[2.87]***	
Age of 25 to35	-0.01	0.039	0.002	0.052	0.002	0.042	
	[0.36]	[0.74]	[0.07]	[1.00]	[0.09]	[0.80]	
Age of 35 to 45	0.017	-0.034	0.022	0.006	0.023	-0.004	
	[0.73]	[0.72]	[0.85]	[0.11]	[0.88]	[0.07]	
Age of 45 to 55	0.036	-0.013	0.051	-0.028	0.051	-0.027	
	[1.61]	[0.28]	[2.10]**	[0.56]	[2.10]**	[0.55]	
Primary education	0.025	0.043	0.029	0.022	0.03	0.027	
	[1.15]	[1.08]	[1.18]	[0.52]	[1.21]	[0.64]	
Lower secondary	0.057	0.052	0.07	-0.01	0.068	0.002	
education	[2.60]***	[1.26]	[2.82]***	[0.22]	[2.76]***	[0.04]	
Upper secondary	0.07	0.098	0.097	0.074	0.095	0.085	
education	[2.82]***	[2.04]**	[3.55]***	[1.41]	[3.49]***	[1.60]	
HH head male	-0.007	-0.051	-0.013	-0.041	-0.013	-0.045	
	[0.69]	[2.48]**	[0.96]	[1.53]	[0.98]	[1.67]*	
HH head married	-0.008	-0.033	-0.006	-0.034	-0.006	-0.033	
	[0.70]	[1.42]	[0.43]	[1.19]	[0.42]	[1.14]	
Household wage	0.028	0.061	0.039	0.053	0.038	0.061	
employment composition	[2.05]**	[2.09]**	[2.58]***	[1.62]	[2.49]**	[1.87]*	
Household	0.023	0.034	0.037	0.096	0.037	0.096	
employment in private sector	[1.55]	[1.10]	[2.02]**	[2.62]***	[2.06]**	[2.60]***	
Being poor commune					0.005	-0.004	
in the province or not					[0.57]	[0.21]	

	ONLY HOUSEHOLD CHARACTERISTICS		STICS	BOTH HOUSEHOLD AND COMMUNE CHARACTERISTICS				
Probability of	BOTH U AND RUR	JRBAN AL AREAS		ONLY RURAL AREA				
Out-Inigration	Economic Long-term Out-migration	Non- Economic Long-term Out-migration	Economic Long-term Out-migration	Non-Economic Long-term Out-migration	Economic Long-term Out-migration	Non-Economic Long-term Out- migration		
Having enterprises, factory, trading village in the circular of 10km					0.003 [0.41]	-0.047 [2.78]***		
Having job-creation program in the commune					-0.003 [0.40]	0.019 [1.24]		
Not good climate for agriculture					-0.004 [0.56]	-0.031 [1.83]*		
Distance to the nearest car way					-0.001 [0.47]	-0.001 [0.54]		
Electricity					0.019 [1.32]	-0.045 [1.34]		
Noteworthy disease					0.007 [1.01]	0.017 [1.18]		
Observations	4008	4008	3100	3100	3100	3100		

Absolute value of z statistics in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

### 4. Migration, Remittances and Income Inequality

#### 4.1. The impact of migration on origin areas

One can distinguish among three different types of impacts of migration on the origin or sending areas, namely departure, diaspora and return impacts (Lucas 2005). *Departure* impacts refer to the impacts of out-migration on labor market, fiscal costs, and externalities in the area of origin. However, the impact generally depends on the composition of the migration flow. Out-migration reduces the overall supply of labor in the area of origin, and particularly in the labor market segments where the migrants were involved, and this tends to raise wages and reduce unemployment levels. But if and to what extent this will happen depends much on barriers to wage flexibility in the market, integration in goods and capital markets, the ease of labor reallocation and the capacity of the remaining labor to obtain the skills of departing migrants, and the time horizon involved. Besides the impacts on the labor market, there are fiscal costs of migration because of a reduction in the tax base. Finally, migration may have an impact on the sending areas through externalities, for instance if a loss of high-skilled migrants deteriorates the quality of management, social climate, educational and health services in the community ("brain drain" effects).

The most important *diaspora* impact of migration is remittances flows as migration typically leads to significant financial flows from the migrants back to the origin areas. Remittances may affect

labor force participation, local demand through a multiplier effect, but also investment in education, land and housing, risk alleviation, inflation, poverty, and intra-family inequality. Remittances will affect the income distribution (income inequality) of an origin area depending on the income classes from which the migrants originate, the pattern of remittances by income classes, and the indirect impacts of migrants on non-migrants (Lucas 2005). Besides remittances, other diaspora effects of migration can be noted as well. For instance, origin communities can benefit from high-skilled migrants, as they become the source of information on various business opportunities and available technologies outside the community, helping to improve regional trade, capital flows and technology transfer.

Migration may also have an impact on origin areas through *return* migration. Returning migrants normally have high saving rates that help them to smooth their retirement period or prolong job search. They might also use their savings for investment in various business fields depending on their experiences, the legal frameworks, credit access and other incentives provided by the area of origin.

In this chapter we will focus primarily on the diaspora impact of migration, particularly remittances and income distribution in origin areas. So far there are virtually no studies on the impact of migration on origin areas in Vietnam. Guest (1998), based on the results of five previous studies on internal migration in Vietnam during 1996-1997, notes that there is little evidence to suggest that rural-tot-rural migration improved opportunities in origin areas. Other studies on the consequences of migration have typically looked at the impact of migration on the migrants themselves or on the destination areas (Do and Nguyen 1998, Le, V.T. 1998, and Nguyen, D.V. 2001). The main reason is that these studies are based on data covering migrants and non-migrants in destination rather than origin areas. For instance, the study of Nguyen (2001) does analyze the probability and amount of remittances sent by migrants, but it is not known how these remittances affect the origin areas.

A number of studies in the international literature have looked at the impact of migration on remittances and income inequality. Recent empirical work suggests that migration and remittances tend to increase household income inequality especially at the initial stage of migration. Stark (1986) and Stark et al. (1988) analyzes the role of remittances in village income inequalities using 1982 household data from two Mexican villages with large out-migration rates. They found that international out-migration had a profound unequalizing impact in the village with few international migrants, but an equalizing effect in the other village with a long history of sending migrants abroad. Taylor (1992), using Mexican data from two household surveys in 1983 and 1989, also finds that initially migration may have a large unequalizing effect before access to migrant labor markets become diffused across households. Adams (1989) measures the impact of international remittances on poverty and income distribution using a 1986/87 household survey in 3 villages in Egypt. He finds that international remittances play a small role in poverty alleviation (percent of households below the poverty line falls from 26.8 to 24.4%) but the income distribution tends to worsen in the presence of migration. Rodriguez (1998) determines the impacts of international migration on the household income distribution in origin areas in The Philippines. Based on the Family Income and Expenditure Survey in 1991, he finds an increase in income inequality, with a 1 percentage increase in remittances raising income inequality by 0.032 percent. Barham and Boucher (1998), using data from Nicaragua in 1991, find that migration and remittances increase income inequality compared to the no-migration counterfactual. Rivera (2005) in a study of remittances and income inequality in Mexico in 2003, finds a negative impact of remittances on household income inequality after decomposing household net income by income source.

In the remainder of this chapter we will analyze the impact of migration on remittances and income inequality in Vietnam using the VHLSS 2002 and 2004. In the next section we provide a descriptive analysis of remittances in Vietnam. Next we estimate the impact of migration on income inequality.

#### 4.2. Descriptive analysis of remittances

The VHLSS 2002 and 2004 ask households whether anyone in the household received money or goods from people overseas ('international remittances') or people in Vietnam ('domestic remittances') who are not household members. If this was the case, the household was also asked about the total (cash and in-kind) value of the remittances. Although this information can be regarded as 'remittances', two caveats should be noted. First, any receipts from non-household members that have never been household members in the past are included as well. Second, any receipts from migrants who have been away for some part of the year and who are counted as household members are not included. In other words, the link between these 'remittances' and migrants who are regarded as household members. Unfortunately there is no way in the VHLSS 2002 and 2004 to distinguish between migrant and non-migrant remittances.

Table 7 reports the average remittances and ratio over expenditure for domestic and international remittances in urban and rural areas in 2002 and 2004. Average remittances are larger in urban than in rural areas, but remittances as a ratio of expenditures are typically higher in rural areas.<sup>62</sup> Domestic remittances dominate international remittances, and total remittances form 12.6-14.9% of total expenditures. Overall we can say that remittances form an important source of household income, both in urban and rural areas.

Tables 8 and 9 report the average remittances and remittances as a ratio of expenditures by income group. Average remittances by income group are very pro-rich (Table 8). While the poorest quintile received only 693,000 VND per household, the richest quintile received 7,239,000 VND per household in 2004. However remittances as a ratio of total expenditures show a very different pattern (Table 9). Domestic remittances are clearly pro-poor while international remittances are pro-rich. This confirms the findings of Stark (1986) and Stark *et al.* (1988) and Taylor (1992) that migration becomes more equalizing with greater access across households. Access to domestic migration has become relatively easy for all households, but access to international migration is still limited and unaffordable for the poorest households because of the necessary employment contract deposit. Total remittances as a ratio of expenditures are both pro-poor and pro-rich – the poorest households benefit disproportionately from domestic remittances and the richest from international remittances.

Type of Remittances	Domestic F	lemittances	Oversea R	emittances	Remit	tances
Area	Urban	Rural	Urban	Rural	Urban	Rural
Remittances (thousa	and VND)					
2002	2,519	1,046	848	640	3,367	1,686
2004	2,636	1,655	1,930	644	4,567	2,299
Ratio over expendit	ure (%)					
2002	8.5	7.8	2.9	4.8	11.3	12.6
2004	7.7	10.7	5.7	4.2	13.4	14.9

Table 7: Average Remittances and Ratio over Expenditure by Type and Area,2002-2004

Note: the averages in the table include also all households with zero remittances

<sup>&</sup>lt;sup>62</sup> Expenditure is viewed as the best proxy for the living standards based from evidence of VHLSS. Expenditure and remittance here are all calculated in real January 2004 prices.

Income	Average Remittances 2002			Average Remittances 2004			
group	Domestic Remittances	Oversea Remittances	Remittances	Domestic Remittances	Oversea Remittances	Remittances	
Poorest	406	71	476	594	99	693	
2	691	82	773	1015	273	1288	
3	751	320	1071	1307	307	1614	
4	1234	1020	2254	2357	734	3091	
Richest	2939	1490	4429	4086	3153	7239	

Table 8: Average Remittances by Type and Income Groups, 2002-2004<sup>63</sup>

*Note*: income groups are defined with respect to per capita household expenditures

Table 3. Hermitiances hallo by Type of Hermitiances and income Groups, 2002-20	Table 9: Remittances Ratio	y Type	of Remittances	and Income	Groups,	, 2002-200
--------------------------------------------------------------------------------	----------------------------	--------	----------------	------------	---------	------------

Income	Domestic Rei	nittance ratio	tio Oversea Remittances ratio		Remittar	nces ratio
group	2002	2004	2002	2004	2002	2004
Poorest	5.35	6.15	0.93	1.02	6.28	7.17
2	7.16	7.96	0.85	2.14	8.00	10.10
3	6.42	8.18	2.74	1.92	9.16	10.10
4	7.95	11.41	6.57	3.55	14.53	14.96
Richest	9.32	9.93	4.72	7.66	14.05	17.60

Note: income groups are defined with respect to per capita household expenditures

The next table reports the characteristics of households by percentage of remittances received in 2004. Most households (61.5%) received remittances between 0 and 10% of total expenditures. Households that received the largest share of their expenditures from remittances are more likely to be rural households. They also tend to have fewer household members, presumably because of migration. Households receiving higher percentages of remittances rely less on wage employment, agricultural activities and, especially, non-agricultural activities. This suggests that non-agricultural activities are economically 'inferior' activities and quickly abandoned with additional income transfers. Households relying most heavily on remittances are more likely to have older, female, unmarried, and less educated household heads. This suggests that remittances may reduce income tend to rely more heavily on remittances, and also suggests that remittances may reduce income inequality. In the next section we will formally analyze how migration and remittances affect household inequality in Vietnam.

<b>Table 10: Characteristics</b>	of Households Receiving	Remittances in 2004
----------------------------------	-------------------------	---------------------

Percentages of total remittances over total expenditure	0%	0-10%	10-25%	25-50%	50-75%	75-100%
% of all households	13.0	61.5	12.5	8.4	2.6	2.0
Urban (%)	26.1	23.6	23.2	19.5	16.8	13.4
Household size (persons)	4.6	4.6	4.1	3.7	2.9	3.0
Per capita expenditure ('000 VND/person)	4787	4834	4831	5268	6463	5465
Income sources (%)						
Wage / Salary paid employment	13.0	17.0	18.0	14.0	8.0	6.0
Agricultural activities	31.3	31.4	25.8	25.9	21.6	16.6

<sup>&</sup>lt;sup>63</sup> Ranking by per capita household expenditure (as the per capita household expenditure is viewed as the best proxy for household living standard from VHLSS)

Percentages of total remittances over total expenditure	0%	0-10%	10-25%	25-50%	50-75%	75-100%
Non-agricultural activities	34.0	27.0	22.0	20.0	16.0	5.0
Other income	20.9	24.3	33.5	40.2	53.9	72.2
Total	100	100	100	100	100	100
Age of household head (years)	46	48	54	57	60	57
Household head is a male (%)	84.1	78.6	72.9	63.4	51.7	58.8
Household head is married (%)	89.1	85.6	76.9	68.8	57.5	64.0
Education of household head (years of education)	7.6	7.4	6.8	6.5	6.0	6.5
Household head has no diploma (%)	25.4	27.3	33.4	36.9	42.0	41.2
Household head has primary education as the highest diploma obtained (%)	25.9	25.8	24.2	25.2	19.8	20.6
Household head has lower-secondary education as the highest diploma obtained (%)	29.8	30.3	27.7	24.3	27.8	19.3
Household head has upper-secondary education as the highest diploma obtained (%)	12.6	12.1	11.9	9.5	7.5	13.1
Household head has tertiary and higher diploma (%)	6.3	4.5	2.7	4.2	2.9	5.9

#### 4.3. Migration, remittances and expenditures

In our analysis of the impact of migration and remittances on household income inequality we follow the framework as developed by Taylor *et al.* (2003). Let the logarithm of total household expenditures be a function of migration (M), remittances from migration (B, and household and community characteristics (Z):

$$\log Y_i = \gamma_0 + \gamma_1 M_i + \gamma_2 R_i + \gamma_3 Z_i + \varepsilon_i^{\gamma}$$
<sup>(1)</sup>

where *i* indicates the household. Total household expenditures depend on migration and remittances for two reasons. First, remittances are a source of income and therefore have a positive impact on expenditures (direct effect). Second, if labor and/or credit markets are imperfect, then the presence of migration and remittances may affect the other production activities of the household (indirect effect). For instance Stark (1991) suggests that the remittances generated by migration may relax credit and risk constraints allowing the household to make a transition from household to commercial production. At the same time if the household faces labor constraints, migration will affect the allocation of labor to the productive activities of the households.

At the same time, remittances are themselves a function of household and community characteristics:

$$R_i = \alpha_0 + \alpha_1 M_i + \alpha_3 Z_i + \varepsilon_i^R \tag{2}$$

Substituting equation (2) into equation (1) gives an equation for the logarithm of total household expenditures as a function of *M* and *Z*:

$$\log Y_i = \beta_0 + \beta_1 M_i + \beta_2 Z_i + \varepsilon_i \tag{3}$$

We will estimate the reduced form equation (3) rather than the structural equations (1)-(2), because we lack a good measure of R. Our measure of remittances includes receipts from non-migrants and excludes receipts from migrants who are regarded as household members, while the variable R in equations (1)-(2) relates to remittances *from migration*. Therefore we estimate the reduced form equation (3) although this implies that we are unable to disentangle the direct and indirect impacts of migration.

The dependent variable of our model (Y) is defined as the logarithm of total household expenditure (in '000 VND) in 2004. As household and commune characteristics (Z) we include the age and age squared (in years), gender (dummy variable for male), highest educational degree obtained (0=no degree, 1=primary school, 2=lower secondary school, 3=upper secondary school, 4=technical worker, 5=professional secondary school, 6=junior college diploma, bachelor's, masters, candidate/doctor) and marital status (dummy for married) of the household head. The age, gender and educational level are included because we expect the human capital of the household head to affect the income and therefore expenditure level of the household. The marital status is included as previous research suggests that married people are more productive.<sup>64</sup> Furthermore we also include household size (number of household members), age composition shares (proportion of household members with age falling into the intervals [0,15], (15,25], (25,35], (35,45], (45,55],  $(55,\infty)$ ), education composition shares (proportion of household members which has not finished 1st grade or never went to school, between 1st and 5th grade, between 6th and 9th grade, and between 10th and 12th grade)65, and the household's total agricultural, syvilcultural and aquacultural land area (ha). We expect households with more resources in terms of people, education or land to be more productive and therefore to have higher expenditures. With respect to the age composition of the household we expect an inverted U-shape – households with many very young or old people (dependents) tend to be less productive and therefore to have lower expenditures. Finally, we include regional dummies among the variables in Z to control for regional differences in income and income-generating opportunities.

VARIABLES	Mean	Min	Max
Log of per capita household expenditure ('000 VND)	9.49	6.78	11.83
HH head age	49.41	15.00	98.00
HH head age squared ('0000)	0.26	0.02	0.96
HH head being married	0.83	0.00	1.00
HH head being a male	0.76	0.00	1.00
HH head education level completed	1.62	0.00	6.00
Household size (persons)	4.40	1.00	20.00
Age of less than 15 (as over household size)	0.27	0.00	0.80
Age of 15 to 25 (as over household size)	0.18	0.00	1.00
Age of 25 to35 (as over household size)	0.13	0.00	1.00
Age of 35 to 45 (as over household size)	0.14	0.00	1.00
Age of 45 to 55 (as over household size)	0.12	0.00	1.00
Age of above 55 (as over household size)	0.16	0.00	1.00
No education (as over household size)	0.13	0.00	1.00
Primary education (as over household size)	0.31	0.00	1.00
Lower secondary education (as over household size)	0.35	0.00	1.00
Upper secondary education (as over household size)	0.21	0.00	1.00
Per capita household agricultural land (hectare)	0.64	0.00	37.87
Short-term out-migration (persons)	0.09	0.00	4.00
Long-term out-migration (persons)	0.44	0.00	10.00
Economic long-term out-migration (persons)	0.08	0.00	4.00
Non-economic long-term out-migration (persons)	0.29	0.00	10.00
No of observations		3935	

**Table 11: Descriptive Statistics of the Model Variables** 

<sup>&</sup>lt;sup>64</sup> This may either reflect the productivity of marriage or a selection effect, as more productive people may be more successful in the marriage market.

<sup>&</sup>lt;sup>65</sup> It is noted that all members that have higher level of education, say bachelor or master or doctoral degree, also record as 12<sup>th</sup> grade.

For the variable *M* in equation (3) we include the number of short-term as well as long-term migrants. In some of the regression specifications we also distinguish among economic (for work reasons) and non-economic (for non-work reasons) long-term migrants.

Table 11 presents the descriptive statistics of the model variables. The only thing we like to note is that the average number of short-term and long-term migrants is respectively 0.09 and 0.44 persons. The average number of economic and non-economic long-term migrants is 0.08 and 0.29 persons.<sup>66</sup>

Table 12 presents the estimation results.<sup>67</sup> In column (1) equation (3) is estimated with OLS and the number of short- and long-term migrants. Household expenditures are higher in households with married heads, with higher education, and with fewer dependents. Surprisingly, female-headed households tend to have higher expenditures than male-headed households. The explanation for this is that female-headed households benefit more from out-migration than male-headed households. If we include interaction terms for the number of migrants and the gender of the household, we find negative coefficients for these interaction terms and the gender dummy becomes smaller and insignificant (not reported in the table).

In terms of migration, we find that households with long-term migrants tend to have slightly higher expenditures but no statistically significant impact is found for short-term migrants. In column (2) the same equation is estimated but with the breakdown between economic and non-economic long-term migrants. This does barely affect the coefficients but we now find the non-intuitive result that non-economic long-term migrants tend to increase household expenditures but economic long-term migrants do not.

The OLS regressions may suffer from endogeneity bias however, as we have seen that migration is a highly selective process. Especially if households that face a negative income shock are more likely to have migrants, we expect that the coefficients on the migrant variables are underestimated.<sup>68</sup> We therefore reestimate equations (1) and (2) with instrumental variables. As instruments we compare the age, gender, marital status and educational level of the household head with the average age, gender, marital status and educational level of the different types of migrants in the region. Intuitively, we assume that households that are similar to migrants in terms of these characteristics are also more likely to migrate. For instance, if many females in one region are migrating because of the demand for female labor in a neighboring region, then we expect that households with relatively more females are also more likely to send away household members. Technically, we define as instruments:

$$x_i^H / (\frac{1}{N_T^{R(i)}} \sum_{j \in S_T^{R(i)}} x_j^M)$$
 for each type of migrant *T* (short-term migrants, long-term migrants)

economic long-term migrants, non-economic long-term migrants), where  $x_i^H$  and  $x_j^M$  are the age, gender, marital status or education of the household head in household *i* and migrant in household *j* respectively,  $S_T^{R(i)}$  is the set of migrants of type *T* in region *R* where household *i* is located (Red River Delta, Northeast, North Central Coast, South Central Coast, Central Highlands, Southeast, Mekong Delta), and  $N_T^{R(i)}$  are the number of migrants of type *T* in region *R* where household *i* is located. Columns (3) and (4) in Table 12 present the estimation results with instrumental variables. Interestingly, although the other coefficients are little affected, we now find strong and significantly positive impacts of migration on household expenditures. Short-term out-migration is estimated to increase household expenditures by 67.7% and long-term out-migration by 38.1%.<sup>69</sup> Column (4) suggests that the impact from long-term migration is particularly high if migrants leave for work (economic reasons) and much less so if they leave for other reasons.

<sup>&</sup>lt;sup>66</sup> This is less than 0.44 (the average number of total long-term migrants) because we do not know the reason for movement of all long-term migrants.

<sup>&</sup>lt;sup>67</sup> The regressions also include regional dummies.

 $<sup>^{68}</sup>$  In other words, there is a negative correlation between the migrant variables and  $\varepsilon_i$ .

<sup>&</sup>lt;sup>69</sup> Exp(0.517)-1=0.677 and exp(0.323)-1=0.381.

	OLS (1)	OLS (2)	IV (3)	IV (4)
HH head age	0.008	0.008*	-0.008	-0.023
	[0.005]	[0.005]	[0.008]	[0.018]
HH head age squared (10 <sup>-4</sup> )	0.026	-0.031	1.264*	2.510*
	[0.415]	[0.415]	[0.668]	[1.480]
HH head being married	0.264***	0.266***	0.262***	0.178**
	[0.029]	[0.029]	[0.036]	[0.079]
HH head being a male	-0.152***	-0.152***	-0.166***	-0.141***
	[0.024]	[0.024]	[0.029]	[0.043]
HH head education completed	0.069***	0.069***	0.071***	0.084***
	[0.007]	[0.007]	[0.008]	[0.014]
Household size	0.143***	0.142***	0.137***	0.164***
	[0.006]	[0.006]	[0.010]	[0.022]
Age of less than 15	0.199***	0.203***	0.320***	0.08
	[0.068]	[0.068]	[0.093]	[0.183]
Age of 15 to 25	0.248***	0.247***	0.131	0.339*
	[0.069]	[0.069]	[0.107]	[0.194]
Age of 25 to35	0.634***	0.630***	0.678***	0.776***
	[0.066]	[0.066]	[0.081]	[0.141]
Age of 35 to 45	0.377***	0.380***	0.403***	0.259*
	[0.060]	[0.060]	[0.073]	[0.137]
Age of 45 to 55	0.253***	0.262***	0.121	-0.094
	[0.052]	[0.052]	[0.083]	[0.219]
Primary education	0.343***	0.343***	0.335***	0.336***
	[0.043]	[0.043]	[0.064]	[0.088]
Lower secondary education	0.653***	0.654***	0.663***	0.618***
	[0.045]	[0.045]	[0.066]	[0.091]
Upper secondary education	1.310***	1.314***	1.366***	1.319***
	[0.055]	[0.056]	[0.076]	[0.115]
Per capita household agricultural land	0.011*	0.011*	0.01	-0.003
	[0.006]	[0.006]	[0.008]	[0.016]
Short-term out-migration	0.01	0.01	0.517	0.393
	[0.021]	[0.021]	[0.331]	[0.457]
Long-term out-migration	0.023***		0.323***	
	[0.008]		[0.116]	
Economic reasons		-0.03		1.852
		[0.022]		[1.158]
Non-economic reasons		0.028***		0.056
		[0.009]		[0.200]
Observations	3935	3935	3935	3935
R-squared	0.54	0.54	0.33	0.33

Table	12: The	Impact of	of Migration	on Household	Expenditures.	2004

Standard errors in brackets. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

#### 4.4. Does migration increase inequality?

Table 9 suggests that domestic remittances are pro-poor, international remittances are pro-rich, and total remittances are both pro-poor and pro-rich (but anti-middle class). However, it would be too quick to conclude that migration increases the relative incomes of the poor and the rich as this depends on what would have happened in the absence of migration. As a first step we could assume that the *counterfactual* household expenditures in the absence of migration would be given by household expenditures minus remittances received. In Table 13 we report the pre- and postremittance expenditure shares of the different quintiles based on this *counterfactual* income. We distinguish among three different counterfactuals, namely (1) no domestic migration (domestic remittances), (2) no international migration (international remittances), and (3) no migration (total remittances). Because each of these counterfactuals leads to different per capita household expenditure rankings, they generate different quintile orderings. The table shows that the expenditure shares of the first quintile increases by 19.6, 44.1 and 82.6% respectively with domestic, international and total remittances. The expenditure shares of the other quintiles fall with remittances. As a consequence, the Gini coefficients of the per capita household expenditure distribution falls after remittances are included in the household expenditures. This finding is in line with Rodriguez (1998, Table 4).

Hence, we may conclude that migration reduces income (expenditure) inequality in Vietnam. However, there are two problems with this conclusion. First, as discussed before, our available measure of remittances is only partly linked with migration, as it includes receipts from non-migrants and excludes receipts from migrants who are regarded as household members. Second, and more fundamentally, it is implausible to assume that (a) migrants would not have contributed anything to household expenditures in case they would have stayed in the household and (b) that the absence of migrants does not affect household expenditures except through remittances. Both assumptions are unlikely to hold. First, in case migrants would remain in the household, they would be available for on- and off-farm income-generating activities and therefore the loss of remittances would be (partly) compensated. Second, in case migrants are absent from the household, household expenditures will not only be affected directly by the availability of remittances (as assumed), but also indirectly in case there are market imperfections. For instance with labor market imperfections, the absence of a migrant worker may decrease farm productivity in case family and hired labor are imperfect substitutes, or with credit market imperfections the availability of remittances may increase off-farm income through investment.

In 2002									
	Expenditure Shares Excluding			Expenditure Shares Including			% Increase in Expenditure Shares		
Income group	Domestic remittances	Oversea remittances	Remittances	Domestic remittances	Oversea remittances	Remittances	Domestic remittances	Oversea remittances	Remittances
Poorest	7.3	6.0	5.2	8.8	8.7	9.6	19.6	44.1	82.6
2	11.4	11.7	11.7	11.4	11.4	11.4	-0.1	-2.3	-2.5
3	15.0	15.2	15.5	14.7	14.7	14.7	-2.2	-3.3	-5.3
4	20.7	21.1	21.2	20.4	20.5	20.3	-1.8	-2.6	-4.4
Richest	45.5	46.0	46.3	44.8	44.7	44.1	-1.6	-2.9	-4.9
Gini coefficient	0.37	0.39	0.40	0.37	0.37	0.37	0.0	-5.1	-7.5
In 2004			•		•			•	
	Expenditure Shares Excluding			Expenditure Shares Including			% Increase in Expenditure Shares		
Income group	Domestic remittances	Oversea remittances	Remittances	Domestic remittances	Oversea remittances	Remittances	Domestic remittances	Oversea remittances	Remittances
Poorest	6.7	6.4	5.2	8.9	8.5	9.9	32.6	32.3	90.3
2	11.4	11.7	11.7	11.4	11.6	11.6	-0.1	-0.2	-1.4
3	15.1	15.4	15.6	15.0	15.1	15.1	-0.7	-2.2	-3.3
4	21.0	21.4	21.5	20.7	20.6	20.5	-1.4	-3.6	-4.3
Richest	45.8	45.1	46.0	44.0	44.2	43.0	-3.8	-2.1	-6.7
Gini coefficient	0.38	0.38	0.40	0.37	0.37	0.37	-2.6	-2.6	-7.5

# Table 13: Pre- and Post-Remittance Expenditure Distribution 2002 - 2004<sup>70</sup>

<sup>&</sup>lt;sup>70</sup> Ranking by "preremittance" per capita household expenditure is defined as per capita household expenditure without domestic, oversea or total remittances (i.e. total household expenditure minus remittances per capita). The preremittance expenditure distribution (i.e. individuals are the same in each of the quintiles) is kept constant when domestic, oversea or total remittances are added to calculate the changes in expenditure shares. Gini coefficients are with respect to *per capita* pre- and post-remittances household expenditures.

Therefore as a second step we compare the distribution of expenditures in 2004 with the distribution of *counterfactual* expenditures in 2004 if no migration had occurred based on our estimated household expenditures model. We first simulate the actual expenditures level for each household:

$$\hat{Y}_i = e^{\hat{\beta}_0 + \hat{\beta}_1 M_i + \hat{\beta}_2 Z_i + \hat{\varepsilon}_i^s} \tag{4}$$

where  $\hat{\varepsilon}_i^s$  is an independent and random draw for each household from the empirical distribution of the estimated error terms,  $\hat{F}(\varepsilon)$ . Next we simulate the *counterfactual* level of household expenditures if there would not have been any recent migration. More specifically, we set *M* equal to zero and adjust *Z* for the fact that the migrant is still present in the household:<sup>71</sup>

$$\hat{Y}_{i}^{C} = e^{\hat{\beta}_{0} + \hat{\beta}_{2} Z_{i}^{C} + \hat{\varepsilon}_{i}^{s}}$$
(5)

where  $Z_i^C$  are the household characteristics in case the migrant(s) would still be present in the household, and  $\hat{\varepsilon}_i^s$  is the same random draw for each household as in (4). The impact of recent migration (including remittances) on the inequality of *per* capita expenditures can then be found by comparing the simulated distributions of  $(\hat{Y}_i/H_i)$  and  $(\hat{Y}_i^C/H_i^C)$ , where  $H_i$  and  $H_i^C$  is the number of household members in the actual and counterfactual situation respectively. However, the simulated distributions of  $(\hat{Y}_i/H_i)$  depend on the series of random draws  $\{\hat{\varepsilon}_i^s\}_{i=1}^N$ , and therefore we repeat the above simulations 10 times and report the mean outcome (Barham and Boucher 1998).<sup>72</sup>

With the estimated coefficients and predicted value of dependent variable, we calculate the total household and per capita expenditure for the counterfactual-model, with no migrants. All household's characteristics are adjusted so that they include the characteristics of the household members who are migrants otherwise. The model chosen is the one with migration variable being the household's number of migrants (column (1) in Table 12).

The total household expenditure in the actual migration model is 20.6% higher than in the counterfactual (no migration) model. In terms of per capita household expenditures the difference is even 29.8%. This suggests that migration does increase economic welfare considerably.

When we compare the distributions of  $(\hat{Y}_i / H_i)$  and  $(\hat{Y}_i^C / H_i^C)$ , we find that the Gini coefficient is 0.04 point *larger* in the actual (migration) model compared to the counterfactual (no migration) model. Hence, unlike in the first step where we compared the pre- and post-remittance household expenditure distributions, we find that migration *increases* the inequality of household expenditures. This finding is exactly similar to that found for the Philippines by Rodriguez (1998), confirming the importance of constructing an adequate counterfactual income level to evaluate the impact of migration on inequality.

Table 14: Household Expenditures in Actual and Counterfactual Model

	Actual model (migration)	Counterfactual model (no migration)
Household expenditure ('000 VND)	18475	15324
Per capita expenditure ('000 VND)	4243	3269
Gini coefficient	0.42	0.38

<sup>&</sup>lt;sup>71</sup> For instance the variables for household size, age and education composition will be affected.

<sup>72</sup> A larger number of simulations does not affect the results.



Figure 1: The Lorenz Curves of Cctual and Counterfactual Per Capita Expenditures

## 5. Discussion and Conclusions

This chapter has used the recent Vietnam Household Living Standard Survey 2004 to analyze the determinants and impacts of migration in Vietnam. Migration is found to be a highly selective process and strongly affected by household and commune characteristics, although differently across types of migration and across urban and rural areas. The evidence suggests that migration pressures will remain strong in the near future as long as the population is young, increasingly educated, and non-farm economic opportunities are lacking. Also the existence of a 'migration hump' for economic long-term migration suggests that migration pressures for long-term migration may be actually increasing with economic development, as more and more of the poor will be able to 'afford' the cost and uncertainty of long-term migration. The presence of non-farm employment opportunities appears to reduce short-term migration but not long-term out-migration for economic reasons, and job-creation programs have been ineffective in reducing migration pressures.

Unlike previous studies on migration in Vietnam, the chapter has also analyzed the impact of migration on household expenditures and inequality in origin areas. Prima facie evidence suggests that migration reduces the inequality in per capita household expenditures in origin areas, as the expenditure shares of the poorest in terms of pre-remittance expenditures increase the most with remittances. However, after constructing a more appropriate counterfactual expenditure level it is actually found that migration increases the Gini coefficient of per capita household expenditures considerably, from 0.38 to 042, in origin areas. This finding is in line with most international studies

that also find that migration tends to increase income inequality. However, it should be noted that international evidence also suggests that the unequalizing impact of migration is reduced over time as access to migrant labor markets becomes diffused across households. Currently, as the pro-richness of international remittances and the evidence for the existence of a migration hump suggest, access to migration labor markets is still limited for the poorest, and this is especially the case for the lucrative international migrant labor markets. Hence, policies improving the access of the poorest to migrant labor markets should be pro-poor.

In terms of impacts the study analyzes the impact of migration on household expenditures and household inequality. Migration is found to have a strong positive impact on household expenditures but increases the Gini coefficient of per capita household expenditures from 0.38 to 0.42 in origin areas compared to the no-migration case.

#### References

- Adams R. 1989. "Worker Remittances and Inequality in Rural Egypt." *Economic Development and Cultural Change* 38(1): 45-71.
- ADB. 2005. "Labor Market Segmentation and Poverty." A report for ADB Making Markets Work for the Poor Project. Hanoi.
- Barham, B., and S. Boucher. 1998. "Migration, remittances and inequality: estimating the net effects of migration on income distribution." *Journal of Development Economics* 55: 307-31.
- Bilsborrow, R. 1996. *Migration, Urbanization, and Development: New Directions and Issues*. United Nations Population Fund and Kluwer Academic Publishers.
- Collier, P., and D. Lal. 1984. "Why poor people get rich: Kenya 1960-79." World Development 12(10): 1007-18.
- Dang, N. A. 1998. "Vai trò của mạng lưới xã hội trong quá trình di cư." ["The Role of Social Networks in the Process of Migration."] In Do Van Hoa, *Chính sách Di dân ở Châu Á* [*Migration policy in Asia*]. Agricultural Publishing House.
- Dang, N. A. 2001. "Rural Labor Out-migration in Vietnam: a Multi-level Analysis." In *Migration in Vietnam-Theoretical Approaches and Evidence From a Survey*. Transport Communication Publishing House.
- Dang, N.A. 2005. Internal Migration: Opportunities and Challenges for the Renovations and Development in Vietnam. Hanoi: The Gioi Publisher.
- Dang, N. A., S. Goldstein, and J. Mc Nally. 1997. "Internal migration and development in Vietnam." International Migration Review 31(2): 312-337.
- Dang, N. A. and Le B. D. 2001. "Women's Migration and Urban Integration in the Context of Doi Moi Economic Renovations." In *Migration in Vietnam-Theoretical Approaches and Evidence From a Survey*. Transport Communication Publishing House.
- Dang, N. A., Le, B. D., and Nguyen, H. V. 2005. "Youth Employment in Vietnam: Characteristics, Determinants and Policy responses." Employment Strategy Papers 2005/09. International Labour Organization, Employment Policies Unit, Geneva.
- Dang, N. A., and Nguyen, T. L. 2006. "Vietnam Migration Survey 2004 Internal Migration and Related Life Course Events". Mimeo. Hanoi: VASS.
- Dang, N. A., C. Tacoli, and Hoang, X.T. 2003. "Migration in Vietnam: A review of information on current trends and patterns, and their policy implications". Paper presented at the Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia, Dhaka, June.
- Dang, N. A., C. Tacoli, and Hoang, X.T. 2004. Stay on the Farm, Weave in the Village, Leave the Home - Livelihood Diversification and Rural-Urban Linkages in the Red River Delta and their Policy Implications. Hanoi: The Gioi Publisher.

- Demery, L. 2002. "Benefit Incidence Analysis." In Tool Kit: Techniques and Tools for Evaluating the Poverty Impact of Economic Policies, chapter 2. World Bank.
- Djamba, Y., S. Goldstein, and A. Goldstein. 1999. "Permanent and Temporary Migration in Vietnam during a Period of Economic Change." *Asia-Pacific Migration Journal* 14 (3): 25-28.
- Doan M. D., N. Henaff, Trinh K. P. 1998. "Di dân nông thôn đô thị ở Việt Nam: Bản chất, mối quan hệ và chính sách quản lý." ["Patterns of Rural-Urban Migration in Vietnam: Nature, relationship and management policy."]. In Do Van Hoa, Chính sách di dân ở Châu Á [Migration policy in Asia]. Agricultural Publishing House.
- Do, M. C., and Nguyen Q. H. 1998. "Rural-rural spontaneous migration: problems and solutions". Paper presented at the International Seminar on Internal Migration: Implications for Migration Policy in Vietnam, UNDP.
- Do, V. H. 1998. "Resettlement in Vietnam: Its Effects on Population and Production". Paper presented in the International Seminar on Internal Migration: Implications for Migration Policy in Vietnam, UNDP.
- Enchautegui, M. 1997. "Welfare Payments and Other Economic Determinants of Female Migration." Journal of Labor Economics 15(3): 529-54.
- GSO. 2006. Human Development Index Profile, Data for National Development Report 2006.
- GSO. 2000. Vietnam Living Standard Survey 1997-1998. Hanoi: Statistical Publishing House.
- GSO and UNDP. 2001. 1999 Population and Housing census: Census Monograph on Internal Migration and Urbanization in Vietnam. Hanoi: Statistical Publishing House.
- GSO and UNFPA [United Nation Population Fund]. 2005. Vietnam Migration Survey 2004: Major Findings. Hanoi: Statistical Publishing House.
- Goldstein, S., Y. Djamba, and A. Goldstein. 2001. "Migration and Occupation Change during Periods of Economic Transition." *Asia-Pacific Migration Journal* 9(1): 65-92.
- Gallup, J. L. 2002. "The Wage Labor Market and Inequality in Vietnam in the 1990s." Policy Research Working Paper 2896. World Bank.
- Guest, P. 1998. "The Dynamics of Internal Migration in Vietnam." UNDP Discussion Paper 1. Hanoi, Vietnam.
- Hanoi Trade Union and ActionAid Vietnam. 2004. Impacts of Economic Integration on Labor of Female Migrants from Red River Delta Rural Areas to Footwear and Textile Companies. Hanoi.
- Harris, J. and M. Todaro. 1970. "Migration, unemployment and development: a two-sector analysis." American Economic Review, 60(1): 126-42.
- Ha P. T., and Ha Q. N. 2000. "Free Rural-Urban Migration Among Female Labour." Hanoi: Women Publisher.
- Katz, E., and O. Stark. 1986. "Labor Migration and Risk Aversion in Less Developed Countries." Journal of Labor Economics 4(1): 134-49.
- Knight, J., and L. Song. 1997. "Chinese Peasant Choices: Farming, Rural Industry or Migration." No.188. Applied Economics Discussion Paper Series. University of Oxford: Institute of Economics and Statistics.
- Lakshmansamy, T. 1990. "Family Survival Strategy and Migration: An Analysis of Returns to Migration." *The Indian Journal of Social Work* 51(3): 473-85.
- Lalonde, R., and R. Topel. 1997. "Economic Impact of International Migration and the Economic Performance of Migrants." In Mark Rosenzweig and Oded Stark, eds., *Handbook of Population and Family Economics* 1B, 799-850, Amsterdam, New York and Oxford: Elsevier Science, North-Holland.
- Le, V. T. 1998. "Dân nhập cư với vấn đề phát triển một đô thị lớn như thành phố Hồ Chí Minh." ["Migrants and Development in a Large City like Ho Chi Minh City"] In Do Van Hoa, Chính sách di dân ở Châu Á [Migration policy in Asia]. Agricultural Publishing House.

- Le, X. B., Cu C. L., Nguyen T. K. D., and Nguyen V. T. 2001. "The Labor Market in Vietnam: Growth, Poverty Reduction and Adjustment to Crisis." Paper presented at the Labour Markets in Transitional Economies in Southeast Asia and Thailand, Phnom Penh, March.
- Lucas, R. 2005. International Migration Regimes and Economic Development. Manuscript.
- MOLISA. 1997. Report on Spontaneous migration survey in Hanoi, Project VIE/95/004. MOLISA: Center for Population and Human Resources Studies. Hanoi.
- National Institute for Urban and Rural Planning. 1996. *Population and Urban Living Environment in Hanoi City*. Ministry of Construction, Project VIE/93/P02. National Political Publishing House.
- Nguyen, D. V. 2001. "Urban Savings and Remittances in Vietnam." In *Migration in Vietnam Theoretical Approaches and Evidence from a Survey*. Transport Communication Publishing House.
- Nguyen, Giang Huong. 2005. "Analysis of Migrants Situation in Hanoi and Ho Chi Minh City." MA thesis. Vietnam-Netherlands Project in Development Economics, Hanoi.
- Nguyen, T. 2002. "A Note on Goods and Labour Mobility in Vietnam." available at <u>http://www.crefa.ecn.ulaval.ca/develop/thang.pdf</u>. Assessed on April 2002.
- Nguyen, T. L. 2001. "Intentions of Temporary Migrants to Stay in Urban Areas. In *Migration in Vietnam-Theoretical Approaches and Evidence From a Survey*. Transport Communication Publishing House.
- Nguyen, T. L. and M. J. White. 2002. "Health status, Urbanward Migration, and Government Policy in Urban Areas in Vietnam." Paper presented at the IUSSP Conference: Southeast Asia's Population in a Changing Asian Context, Siam City Hotel, Bangkok, Thailand, June.
- Rivera, J.J.M. 2005. "The impact of migration and remittances on distribution on distribution and sources income: the Mexican rural case". Paper UN/POP/MIG/2005/06 presented at the United Nations Expert Group Meeting on International Migration and Development. Population Division, Department of Economics and Social Affairs, United Nations Secretariat. New York.
- Rodriguez, E. 1998. "International Migration and Income Distribution in the Philippines." *Economic Development and Cultural Change* 329-50.
- Stark, O. 1980. "On the Role of Urban-Rural Remittances in Rural Development." Journal of Development Studies 16(3): 369-74.
- Stark, O. 1986. "Remittances and Inequality." The Economic Journal 96: 722-40.
- Stark, O. 1986. "Migration, Remittances and Inequality." Journal of Development Economics 28: 309-22.
- Stark, O. (1991): The Migration of Labour. Cambridge, Mass.: Harvard University Press.
- Stark, O., E. Taylor, and S. Yitzhaki (1988): "Migration, Remittances and Inequality. A
- Sensitivity Analysis Using the Extended Gini Index", *Journal of Development Economics* 28: 309-22.
- Stark, O., and J. Taylor. 1991. "Migration incentives, migration types: The role of relative deprivation." *The Economic Journal* 101: 1163-78.
- Taylor, J. 1992. "Remittances and Inequality Reconsidered: Direct, Indirect, and Intertemporal Effects." *Journal of Policy Modeling* 14(2): 187-2008.
- Taylor, J., S. Rozelle, and A. de Brauw. 2003. "Migration and Incomes in Source Economies: A New Economics of Migration: Perspective from China." *Economic Development and Cultural Change* 52: 75-101.
- Thadani, V., and M. Todaro. 1984. "Female Migration: A Conceptual Framework." In J. Fawcett et al., eds., *Women in the Cities of Asia. Migration and Urban Adaptation*, Boulder: Westview Press.
- Todaro, M. 1969. "A model of labor migration and urban unemployment in less developed countries." *American Economic Review* 59(1): 138-48.
- Trinh, K. P. 1998. "Characteristics of Rural-Urban Migration in Vietnam and Policies to Control It." Paper presented at the International Seminar on Internal Migration: Implications for Migration Policy in Vietnam, UNDP.
- Truong, S. A., P. Gubry, Vu T. H., and J. Huguet. 1996. "Migration and employment in Ho Chi Minh City." *Asia-Pacific Population Journal* 11(2): 3-22.
- UNDP and GSO. 2001. *Living Standards during an Economic Boom Vietnam 1993-1998*, Hanoi: Statistical Publishing House.
- VASS. 2006. "Q-Squared (Q<sup>2</sup>): Combining Qualitative and Quantitative Approaches in Poverty Analysis in Vietnam." Mimeo.
- Vu, T. H., and L. Monteiro. 2001. "The Health of Migrants in the Reforming Vietnamese Economy." In Migration in Vietnam-Theoretical Approaches and Evidence from a Survey. Transport Communication Publishing House.
- White, M. J., Djamba, Y. K., and Dang, N. A. 2000. "Implications of Economic Reform and Spatial Mobility on Fertility in Vietnam." Working Paper 2000-01. Population studies and Training Center.
- Winkels, A., and W. Adger. 2002. "Sustainable Livelihoods and Migration in Vietnam: The Importance of Social Capital as Access to Resources". Paper presented at the International Symposium Sustaining Food Security and Managing Natural Resources in Southeast Asia - Challenges for the 21<sup>st</sup> Century, Chiang Mai, Thailand, January.
- Wood, A. 1991. "North-South trade and female labour in manufacturing: an asymmetry." *The Journal* of *Development Studies* 27(2): 168-89.
- WB and MOT [Ministry of Transport]. 2006. "Social Assessment Report." Mekong Transport Infrastructure Development Project (MTIDP). Hanoi.
- WWF [World Wild Fund]. 1999. "Human Migration and Resource Utilization." Research Project of Population Dynamics and Resource Utilization in Buffer Zone of Yokdon National Park and the Surrounding Region Daklak Province.

# PART II: POLICY AND POVERTY REDUCTION

4

# Better Than Its Reputation? The Incidence Of Social Transfers And Education Fee Exemptions In Vietnam

Henrik Hansen and Le Dang Trung

Together with high economic growth rate over the last few years, redistribution of public funds has been a crucial topic for both researchers and policy makers. Since 1997 there have been at least six analyses on the incidence of public transfer programs which make use of the 1992/93 and 1997/98 VLSSs. These analyses mainly do not recognize the success of the Vietnamese government's poverty-targeted redistributive schemes when concluding that the social transfer system is regressive, and that it does not provide effective social protection to the majority of the poor. With the availability of new data, namely, the 2002 and 2004 VHLSSs, the incidence analysis can be updated. We seek answers to the following questions: what is the incidence of the redistribution programs? Are they targeted to the poor? What are the roles of these programs in poverty dynamics? In doing so, we challenge many of the previous studies. We raise concerns about (1) the choice of the counterfactual welfare measure in van de Walle's analyses and (2) the appropriateness of the household survey data for doing the analyses.

In estimating the counterfactual welfare measure which is the marginal propensity to consume out of transfer (*mpct*) based on the 2004 VHLSS, we notice that the *mpct* seems not to be precisely defined. Therefore, even though our point estimate is one, we cannot reject the hypothesis that the true value is only one-half as found in van de Walle (2004a). Instead, we come to conclude that the *mpct* could lie anywhere from about 0.1 to more than 1. Thus, the social transfer incidence results should be interpreted with great caution.

In the analysis, we look at two types of transfers: social insurance and social protection, and two types of education fee exemptions: tuition fee exemptions and school contribution fee exemption. The results paint a somewhat different picture of the distribution of social transfers than the studies based on the 1992/93 and 1997/98 household surveys. Fist of all, social transfers are not regressive in 2004. Second, the transfers had a sizeable impact on the level of poverty in both 2002 and 2004 and, furthermore, provided protection against falling into poverty in that period.

#### 1. Introduction

Vietnam is in many ways a remarkable country. At present, most observers praise the macroeconomic

performance of Vietnam during the last twenty years. The economy has been growing rapidly and the growth has been accompanied by a reduction in poverty that appears second to none. From 1993 to 2004 the average annual growth rate in per capita GDP was 5.9 percent while the poverty headcount rate fell from 58 to 20 percent implying a growth elasticity of 1.6 over the period. Hence, growth has been high and inclusive in Vietnam.

One reason for the inclusiveness is an active policy of redistribution of public funds across provinces in Vietnam. The extent of redistribution is illustrated in Figure 1, in which net transfers per capita (within provinces) to and from the central government in 2003 is plotted against province level poverty headcount rates in 2002. It is clear that the growth centers, with low poverty rates, transfer large amounts to the central government, which subsequently supports the poorer provinces, mainly in the Central Highlands and the Northern Uplands.

Despite this large inter-provincial redistribution, there is a general perception of poorly targeted redistributive transfers at the household level. That is, direct government support to households in the form of social transfers has been found to be either regressive or only mildly progressive in studies of the social safety net in Vietnam. Preston (1999) provides a thorough description of the institutional background and the public expenditure on social transfers in the 1990s. He concludes that Vietnam in the 1990s was divided into a minority group in the formal sector that had a well developed social security system and a majority, mostly rural, group with little, and clearly insufficient, social protection. This conclusion seems to be representative of all studies of the social transfer system in Vietnam.



Figure 1: Provincial Poverty Rates in 2002 and Net Government Transfers To and From Provinces in 2003

*Notes:* Net transfers are calculated as the difference between the actual spending of a province (both current and capital expenditures) and the tax revenue raised on its territory. *Source:* World Bank (2004). In another study, by Prescott (1997), the incidence of social insurance transfers and social protection transfers are analyzed at the household level using the household survey 1992/93 VLSS. Prescott finds that social transfers, as such, are not pro-poor. The social insurance system (covering public sector workers) is regressive as it mainly benefits the more well-off households both in terms of the value of transfers and as a share of total consumption expenditure. Moreover, the coverage of the social protection program, which is mildly progressive, is very low and the size of the transfers is small compared to the social insurance transfers. Prescott concludes that a significant reallocation of the budget from social insurance to social protection is required to improve the poverty orientation in social transfers.

Household level incidence analyses, based on the 1992/93 and 1997/98 VLSSs are also presented in the most comprehensive studies of social transfers in Vietnam by van de Walle (2004a, b). van de Walle first addresses the issue of behavioral responses to social transfers by estimating the marginal propensity to consume out of transfer income. The result of the econometric analysis is that households benefiting from social transfers only increase consumption expenditure by half of the transfer. This leads van de Walle to define a counterfactual consumption measure—in which half of the value of transfers is subtracted from actual per capita consumption—which she subsequently uses in the incidence analyses. Based on a detailed analysis of several transfers van de Walle concludes that transfers to households are negligible and coverage is weak. Moreover, although transfers in 1998 are progressive they are not particularly well targeted as the poor receive less in absolute amounts than the non-poor. Furthermore, in an analysis of the dynamic incidence of transfers van de Walle comes to the conclusion that social transfers had, at most, a very small impact on the large decline in poverty from 58 percent in 1993 to 37 percent in 1998. Also, the transfers did not provide protection against falling into poverty between 1993 and 1998.

In addition to the four studies above, the World Bank (2002) and Justino (2005) briefly report some of van de Walle's results. Both studies conclude that the social transfer system is regressive, and that it does not provide effective social protection to the majority of the poor.

In this chapter we provide an update of the incidence analysis of social transfers to 2004 using the household surveys 2002 and 2004 VHLSS. In doing this we also challenge many of the conclusions in the previous studies. Our dispute is based on three arguments. First of all, the choice of counterfactual welfare measure, where van de Walle (2004a, b) use observed per capita consumption net of half the per capita transfer because households only spend half of the social transfers on consumption goods, has a decisive impact on the distribution of transfers as the social insurance benefits are quite large. In our econometric analysis we find that households seem to fully consume the transfers and this leads us to consider a counterfactual welfare measure defined as per capita consumption net of (all of) the per capita transfer. Comparing the results of using the two counterfactual welfare measures we find that the distribution of social insurance transfers changes from a U-shape, in which the absolute value of the transfers mostly benefit the poorest and the richest quintiles, to a regressive distribution. Hence, conclusions about the incidence of the social insurance transfers depend on the choice of welfare measure. The more recent data supports the counterfactual leading to a reasonably progressive U-shaped distribution of social insurance transfers. However, we cannot reject the alternative.

The choice of counterfactual also impacts on the dynamic incidence analysis. In contrast to van de Walle's finding of no protection from falling into poverty between 1993 and 1998, our results indicate that the social transfers had sizable effects on the poverty levels in 2002 and 2004 and, furthermore, provided significant protection against falling into poverty.

Second, while we agree with the critique of the social transfer system in terms of too limited resources for social protection transfers, resources have increased slightly since 1998 to reach 16 percent of the poorest individuals in 2004. Yet, this is still a low incidence. However, we find it difficult to do an in-depth analysis of the incidence of the social protection transfers. The problem is

two-fold: The household surveys are designed to provide a cost-effective overall picture of 'typical' households in Vietnam in the sense that the samples are fairly small and the stratification is purely geographical. This means that the relatively small groups of extremely poor, as defined by the Government of Vietnam, are not well represented in the surveys. This may lead to a downward bias in the estimated transfers to the poorest group. Furthermore, the social protection program includes transfers to individuals with a historical merit (*i.e.*, people with a revolutionary merit and their relatives). Such beneficiaries receive transfers independently of their level of income and consumption. As we cannot distinguish between means tested transfers and 'war compensation transfers' in the social protection program it is impossible to make firm conclusions about the targeting of the program.

Finally, the social transfer system is only one part of the direct government support to households. Another, substantial, support is provided as exemptions from education-related user fees. Based on the data in the 2004 VHLSS it is possible to estimate the total value of education fee exemptions granted to households. The total value of education fee exemptions is three times larger than the total expenditures on social protection. Hence, to get a more complete picture of the redistributive transfer policies in Vietnam one should include the incidence of education fee exemptions.<sup>73</sup> Our incidence analysis of the fee exemptions shows that this support is progressive and fairly well targeted, considering that the fee exemption programs use both direct and categorical targeting.

In the Vietnamese household surveys it is possible to get information on the reason for education fee exemptions at the household level. We use this information to define a set of directly poverty oriented education fee exemptions by separating out fee exemptions granted for other reasons, including 'war compensation grants'. The division into directly poverty oriented fee exemptions and other fee exemptions shows that the former are clearly pro-poor in terms of coverage. The main problem with the program is that the government cannot control the size of the subsidy, and there is a clear tendency for richer households to pay higher education fees, thereby benefiting from larger subsidies when they are exempted from the fees.

In an analysis of the incidence of the joint government support system, including both social transfers and education fee exemptions, the results show that about one-quarter of the Vietnamese people are affected by the joint support program. Of these, roughly 19 million people, some 37 percent (7 million) are in the poorest quintile of the counterfactual per capita consumption expenditure distribution. Moreover, 75 percent (14 million) of the beneficiaries live in the rural areas of Vietnam. Hence, if one considers a social safety net as a system that includes both direct money transfers (increasing household income) and fee exemptions (lowering household expenditure) then the above conclusion of a division into an urban minority of well secured people and a large rural majority of poorly covered people is no longer evident.

A third part of the direct government support to households is the set of National Targeted Programs (NTPs), of which the Hunger Eradication and Poverty Reduction (HEPR) program is the best known. The NTPs provide a wide range of support options for the poor, however, not much in terms of direct transfers apart from education fee exemptions, which we cover. Therefore, we do not look at the targeting of the NTPs in the present chapter. Instead, we refer to Nguyen (1999), Nguyen (2003), Shaffer and Nguyen (2004) and UNDP-MOLISA (2004) for analyses of targeting and impact of the HEPR.

The chapter is structured as follows. Section 2 briefly describes the institutional setting and the total expenditures for social transfers and education fee exemptions. In Section 3, we estimate the marginal propensity to consume out of transfer income and education fee exemptions. Next, in

<sup>&</sup>lt;sup>73</sup> van de Walle (2004b) looks at the coverage of education fee exemptions. But it is not possible to estimate the value of the fee exemption in the 1997/98 VLSS, therefore, she cannot carry out a complete incidence analysis.

Section 4 we analyze the distribution of social transfers and education fee exemptions across individual welfare levels using counterfactual measures of per capita consumption expenditure as the welfare indicator. The dynamic incidence of social transfers from 2002 to 2004 is analyzed in Section 4. Finally, Section 5 concludes and provides a few policy recommendations.

## 2. A Public Expenditure View on Social Transfers and Education Fee Exemption

In this section we first describe the social insurance and social protection systems to establish an overall view of the public expenditures on these programs, the institutions and the intended beneficiaries. Second, we provide a brief description of the education system focusing on direct subsidies to households in the form of tuition fee exemptions and school contribution fee exemptions (in brief 'education fee exemptions').

#### 2.1 Social Transfers

In 2002 the Government of Vietnam spent 13.2 trillion VND of the state budget on expenditures classified as social insurance. This amounts to 10 percent of the total budget. As most of social insurance expenditure are current expenditure (12.8 trillion VND), the social insurance expenditure took up 16 percent of the total current expenditure budget that year. This makes social insurance the second largest current expenditure component in the state budget only exceeded by education and training expenditure, which constitutes 21 percent of the current expenditure budget.<sup>74</sup>

The largest single component of the social insurance expenditure is the social insurance system, which provides cash transfers for old-age in the form of pension, early retirement and survivorship. In addition, the scheme also covers maternity, sickness related to work accidents and occupational diseases. The social insurance system started as a government retirement program already in 1947. Until 1995 the system was fully state-funded and included only public servants, employees of state owned enterprises (SOEs), the armed forces and members of the Communist Party's organizations. In 1995 it was reformed to a modern social insurance system by the establishment of the Vietnam Social Insurance (VSI) program. The VSI is an independent entity directly under the Prime Minister, which is managed by the Ministry for Labor, Invalids and Social Affairs (MOLISA) and funded by pay-roll taxes with contributions from both employers and employees (15% and 5% of basic salary, respectively) and by contributions from the state budget.

With the establishment of the VSI the coverage was extended from the narrow state and party related employees to compulsory membership of all employees of non-SOEs with 10 workers or more and all employees of foreign invested enterprises and foreign organizations. Subsequently, the program has been enlarged to offer voluntary membership to all formal sector workers not included in the mandatory membership group. However, the total coverage of the VSI is still small. At the end of the 1990s and the beginning of the 2000s the VSI covered only some 14 percent of the labor force (MOLISA, 1999 and World Bank, 2004).

During a transition period in which the VSI is building up resources individuals who retired before 1 January 1995 are still covered directly from the state budget. In 2004, about one million individuals received monthly pensions directly from the state budget while almost 322 thousand individuals were covered by the VSI. But, as expected, the recent statistics show a slow decrease in the number of people covered by the state budget and an increase in the number of people covered by the VSI.

<sup>&</sup>lt;sup>74</sup> Expenditure classified under "Transport, Storage and Communication" makes up a larger fraction of the total expenditure (14 percent). However, this is almost entirely capital expenditure.

Pensions from the new social insurance system are provided on a pay-as-you-go basis and the size of monthly pension benefit is directly related to a weighted average of the monthly salary and the number of years an employee has been contributing to the VSI program. For employees contributing up to 15 years the monthly pension benefit is 45 percent of the average monthly salary. The monthly benefit gradually increases to a maximum of 75 percent of the average monthly salary after 30 years of contributions to the program.<sup>75</sup> Currently the pension age is 55 for women and 60 years for men. Employees in some special hazardous occupations may retire five years before the general retirement ages.

Clearly, the social insurance program is not a poverty oriented welfare program as such; it is a standard social insurance scheme that mainly provides old-age, disability and survivors' benefits to its members. Hence, when looking at redistributive government transfers the main problem in relation to this program appears to be the small coverage of workers combined with the large contribution from the state budget.

In addition to the social insurance program, Vietnam has a social protection program transferring social allowances based on specific eligibility criteria. The social protection program is organized in several social guarantee funds (SGFs) of which the three most important are "the Social Guarantee Fund for War Affected Groups", "the Social Guarantee Fund for Regular Relief" and "the Contingency Fund for Pre-Harvest Starvation and Disaster Relief". The specification of beneficiaries and the benefit levels for those covered by the SGF for War Affected Groups are specified in Decree 210/2004/ND-CP while beneficiaries and benefit levels for the two relief funds are specified in Decree 07/2000/ND-CP. Extracts from the two decrees are given in Appendix 1.

The Social Guarantee Fund for War Affected Groups (or the Social Guarantee Fund for Veterans, War Invalids and Others Who Have Contributed to National Liberation Struggles) supports war veterans and relatives of people that were killed or disabled in the Revolutionary cause (see Appendix 1).<sup>76</sup> Thus, a means testing of the beneficiaries from this fund does not involve the level of welfare in terms of consumption or income. Hence, in some sense it would seem more appropriate if the transfers from this fund were classified as part of a 'war compensation program' rather that a part of the social protection system.

In contrast, The Social Guarantee Fund for Regular Relief provides monthly benefit support to three groups of individuals that are regarded as being extremely poor and unable to support themselves. The three groups are the lonely elderly, orphaned children having no support from relatives, and the poor disabled people (Appendix 1 and MOLISA, 1999). The eligible beneficiaries are identified by local authorities and while a minimum allowance for each group is specified by the government the local authorities are encouraged to increase the allowances if sources are available. This means that actual transfers to the beneficiaries vary somewhat across provinces, with higher benefit levels in relatively well-off provinces. MOLISA (1999) reports that more than one million people were eligible for regular social allowances in the late 1990s but only one in five received assistance. Due to lack of updated data, it is not clear if this rate of coverage has improved since 1999.

Besides the regular allowances the regular relief fund also finances centers for rehabilitation of social evils groups; drug addicts, prostitutes, and HIV/Aids positive. As for the recurrent regular relief, the coverage in the late 1990s was low and it is unclear if it has increased in recent years.

<sup>&</sup>lt;sup>75</sup> The calculation of the average monthly salary depends on the employment situation as there is a distinction between state and non-state employment. For an employee who has been in the state salary system throughout, the average monthly wage is based on the last 5 years of employment. Otherwise, the average monthly salary is based on the whole period in which the employee has been contributing to the pension scheme. Needless to say this discriminates private sector employees as the calculations do not appear to take account of inflation.

<sup>&</sup>lt;sup>76</sup> Relatives include widows, parents, parents-in-law, children, and even other dependent individuals who are not family related.

The third fund, the Contingency Fund for Pre-Harvest Starvation and Disaster Relief, has a completely different purpose compared to the two other funds as it provides short-term support to victims of natural disasters and to households that may suffer extreme hunger between harvests. The hunger relief is given in the form of food aid or cash to buy food. The benefit amount is determined by the local authorities and may vary. In 1999 the support amounted to 25-30 thousand VND per person per month (MOLISA, 1999). The emergency relief to households in disaster struck areas is given for injury relief, funeral fees and assistance for repairing or restoring seriously damaged houses. Also for the disaster relief, the benefit level is determined by local authorities (Appendix 1). In 1999 the official level of support was about one million VND per household for restoring houses; one million VND for funeral assistance; and 200-500 thousand VND for injury relief (MOLISA, 1999).

Preston (1999) points to an important practical problem in covering both disaster relief and preharvest starvation in one fund. Due to the unforeseeable nature of natural disasters, the support to poor households suffering from pre-harvest starvation varies substantially across regions and years as this support is used as a buffer with large support in years without natural disasters and low support in years in which a disaster occurs. In terms of redistribution and targeting of the poor this means that the incidence of these transfers may well vary greatly from one year to the next; at least within the lower income quintiles.

Because of lack of updated data on the social insurance system and the social protection program we use the 2002 and 2004 household surveys (VHLSS) to estimate the total expenditures on social transfers. Information on both social insurance and social protection is available at the household level. However, for social insurance transfers we cannot separate state payments from VSI payments and, likewise, we cannot identify benefits from the individual funds in the social protection program. To facilitate comparison the transfers in both years are measured in real January 2004 prices.<sup>77</sup>

In both surveys, a few households report extraordinary large social transfers. We identify such outlying observations using an iterative outlier detection algorithm suggested in Hadi (1992, 1994) and subsequently omit these observations from the estimation of total social transfers.<sup>78</sup> This strategy of removing the extreme observations will be followed throughout. Needless to say, omitting the extreme observations will have a large impact on the distributional results that are the core of the analysis in this chapter. But we are convinced that these observations are results of misunderstandings or misreporting.

Table 1 reports the total benefits received by the households using the survey sampling weights to estimate the total expenditure on social insurance and social protection. The estimated total social transfers in 2002 amounts to 11.7 trillion VND (10.8 trillion VND in 2002 prices).<sup>79</sup> Almost 85 percent of the total expenditures are social insurance transfers. The social insurance transfers have a heavy urban bias as 44 percent went to individuals in urban households while these individuals only make up 23

<sup>&</sup>lt;sup>77</sup> Social insurance transfers are from question m5d1c2t103 in VHLSS 2002 and from question m4d1c2\_3 in VHLSS 2004. Social protection transfers are from questions m5d1c2t104 and m4d1c2\_4, respectively. The transfers in 2002 are first deflated to real January 2002 prices using regional and monthly deflators provided with the 2002 VHLSS data. Subsequently the 2002 transfers are inflated to January 2004 prices using the consumer price index from GSO.

<sup>&</sup>lt;sup>78</sup> The outliers are identified using the algorithm developed in Hadi (1992, 1994) with a one percent significance level for outlier cutoff. Only non-zero values of the transfers are considered in the algorithm.

<sup>&</sup>lt;sup>79</sup> The estimated total expenditure is 1.9 trillion VND smaller than the current expenditure on social insurance as reported in the State Budget (SRV, 2005). Given (i) the sampling uncertainty of the estimate, (ii) the fact that the expenditure estimated in the survey does not include administrative costs, and (iii) that transfers in the survey period are not exactly covering the fiscal year 2002, we do not find this discrepancy to be too alarming. However, it is important to keep in mind that the discrepancy may be caused by a systematic misrepresentation or omission of some beneficiaries in the household surveys.

percent of the population. In contrast, of the 1.8 trillion VND received as social protection transfers, about 1.5 trillion VND were benefiting households in the rural areas. Both of these biases are expected. The social insurance system is focused on former public employees and former employees in large enterprises, both of which are mainly located in urban areas. Moreover, as the regular relief fund is focused on natural disasters and pre-harvest starvation, the main beneficiaries from this fund will be in the rural areas.

		2002 VHLSS			2004 VHLSS		
	Rural	Urban	Total	Rural	Urban	Total	
Social insurance	5524.5	4397.9	9922.5	7601.0	7538.0	15139.0	
	(55.7)	(44.3)	(100.0)	(50.2)	(49.8)	(100.0)	
Social protection	1484.9	330.1	1814.9	1819.6	321.2	2140.7	
	(81.8)	(18.2)	(100.0)	(85.0)	(15.0)	(100.0)	
Memo item:							
Population structure (percent)	72.5	27.5	100.0	70.2	29.8	100.0	

Table 1: Social Transfers from 2002 and 2004 VHLSS

Note: All monetary values are in billion VND, January 2004 prices. Numbers in parentheses are percentage shares of the total expenditures.

Source: Own calculations based on 2002 and 2004 VHLSS

Social insurance increased by more than 50 percent in real terms from 2002 to 2004. This is a large increase even considering the rapid growth of the Vietnamese economy.<sup>80</sup> The bulk of the increase in social insurance transfers went to households in the urban areas, thereby increasing the urban bias in these transfers. Part of the increase in the urban area may be explained by the urbanization in the period, illustrated by the change in the composition of the population in the bottom part of Table 1.

The increase in social protection benefits was more modest (18 percent) resulting in a decrease in the share of social protection in total social transfers from 15 to 12 percent. Hence, in relative terms, the direct poverty orientation in the social expenditures has decreased, even though the expenditures have increased in real terms.

#### 2.2. The Education System and Rules for Education Fee Exemption

The education sector, as seen from a public expenditure perspective, has been described recently in SRV (2005) while other aspects of the sector, in particular, the long run policies, are given in SRV (2003). Here, we therefore focus on issues of relevance for understanding the system of education fee exemptions, and refer to the SRV (2003, 2005) reports for detailed descriptions of education policies, expenditures and service delivery issues.

The school system in Vietnam has four basic levels: pre-primary, primary, lower secondary and upper secondary. Primary and lower secondary schooling can be followed by vocational training while higher education in the form of college and university degrees is possible after upper secondary school. The overall responsibility for the education sector lies in Ministry of Education and Training (MOET) but the system is decentralized such that communes and districts manage the first three levels of schooling (pre-primary to lower secondary) and provinces manage upper secondary schools. Vocational training is managed by MOLISA while higher education is directly under MOET. The decentralization includes state budget management whereby the relevant local authorities are responsible for budget allocations to finance local educational activities. In 2002, some 98 percent of education expenditures (excluding vocational training) were administered by local authorities (SRV, 2005).

<sup>&</sup>lt;sup>80</sup> The growth rate in real GDP was 15.6 percent from 2002 to 2004. (Based on data from World Bank, 2005).

Sub-sector	1993	1998	2002
Primary			
State subsidy	45	55	73
Contributions and direct expenditures by parents	55	45	27
Lower secondary			
State subsidy	34	62	59
Contributions and direct expenditures by parents	66	38	41
Upper secondary			
State subsidy	40	47	52
Contributions and direct expenditures by parents	60	53	48
Source: SRV (2005).			

Table 2: Funding Sources of Education Expenditures (percent)

Total education expenses are partly covered by community contributions (user fees). These contributions are regarded as state budget revenues. They are collected and retained by the local educational institutions to finance the local educational activities. The main community contributions are tuition fees, examination fees, and contributions for school construction.

The distribution of total education expenditure across public and community contributions is shown in Table 2. The share of contributions is increasing with the level of schooling. One reason for the low share of user payment at the primary level is that almost all households are exempted from tuition fees at this level, though parents still have to pay other fees and contributions. Moreover, it is interesting to note that the share of community contributions in total expenditure has fallen steadily over time at all levels of schooling.

The importance of community contributions in the funding of the education sector has created a tension in the education policies as it is recognized that user fees is one of the obstacles for increasing school participation among the poorest groups. Therefore, the Government has expressed commitment to provide free primary education for all by 2015 and, according to the plans in SRV (2003), the community contributions to lower secondary education will be kept almost constant in real terms from 2003 to 2015 resulting in a sharp decline in the share of community contributions in total expenditures.

More important for the present study, the Government has established policies for tuition fee and school contribution exemptions and reductions. The policies are typically managed at the local level through targeted programs. Several groups of students are intended to be exempted from paying tuition fees and other contributions.<sup>81</sup> The list of students includes war veterans and children of people who were killed or disabled in the Revolutionary cause; minority people; people living in extremely difficult areas; and invalids.<sup>82</sup>

The list of intended beneficiaries shows that education fee exemption is based on categorical targeting as poor households are targeted indirectly through the high incidence of poverty among ethnic minority groups and the high incidence of poverty in extremely difficult areas. Moreover, as for social protection, education fee exemption can be based on historical merits rather than a means based assessment.

A more direct poverty orientation in is stated in the National Plan for Education for All where the Government commits to grant "exemption of user cost for students from poor families and other disabled groups" at the lower secondary level (SRV, 2003, p. 63). However, it is not clear when and how this commitment will be introduced and administered.

<sup>&</sup>lt;sup>81</sup> Some education fee exemptions are given in Decree No. 88/2001/ND-CP of November 22, 2001.

<sup>&</sup>lt;sup>82</sup> Education fee exemption is also granted to poor students in non-formal education (literacy and complementary programs for out-of-school youth and young adults with low educational skills). However, the non-formal education program is small and not covered in the present study.

To our knowledge there are no official statistics on the total value of education fee exemptions. Therefore, we rely on the information in the 2004 VHLSS to estimate such numbers. In the survey, each individual who participated in the education system during the last twelve months before the interview is asked about the amount of school fees and the amount of other contributions the household paid regardless of whether or not the individual was granted education fee exemption. A second question asks if the individual was exempted from paying education fees and other school contributions and if so there is a follow-up question about the rate of reduction. The rate of education fee exemption varies in the range from 0 to 100 percent.

As for the social transfers, some households report rather extreme values of education fee exemptions. Therefore, we also apply the Hadi (1992, 1994) outlier detection algorithm for education fee exemptions and compute totals for the sample excluding the outliers.

Table 3 shows the estimated value of education fee exemptions in 2004 by type of fee and by rural/urban area. Tuition fee exemptions and school contribution fee exemptions both exceed total expenditure on social protection whereby the education fee exemption program is more than three times the size of the social protection program in terms of resources transferred to the households. This is the reason why we argue that this program should be evaluated alongside social transfers in assessments of redistributive transfers in Vietnam.<sup>83</sup>

Both tuition fee and contribution fee exemptions have a small rural bias as the rural share of fee exemptions is higher than the population share. The bias is present at all levels of schooling, the only exception being the unspecified tuition fee exemptions.

	Ru	Rural		Urban		
		Percent of		Percent of		
	Exemption	row total	Exemption	row total		
Tuition fee exemption						
Primary	71.9	(89.2)	8.7	(10.8)	80.6	
Lower secondary	1,075.8	(78.1)	302.4	(21.9)	1,378.2	
Upper secondary	833.0	(73.4)	301.7	(26.6)	1,134.8	
Unspecified	233.9	(61.6)	145.6	(38.4)	379.5	
Total	2,214.6	(74.5)	758.5	(25.5)	2,973.0	
Contribution fee exemption						
Primary	1,538.2	(88.8)	193.4	(11.2)	1,731.6	
Lower secondary	1,109.1	(83.9)	212.4	(16.1)	1,321.5	
Upper secondary	557.7	(79.7)	141.8	(20.3)	699.5	
Unspecified	231.8	(91.6)	21.2	(8.4)	253.0	
Total	3,436.8	(85.8)	568.8	(14.2)	4,005.6	

Table 3: Total Value of Education Fee Exemptions from 2004 VHLSS

Note: All monetary values are in billion VND, January 2004 prices.

Source: Own calculations based on 2004 VHLSS.

<sup>&</sup>lt;sup>83</sup> Some of the previous studies of social transfers in Vietnam (van de Walle, 2004b; the World Bank, 2002 and Justino, 2005) include scholarships in the analyses of social transfers, noting that the number of scholarships is very small. This is not surprising as scholarships are typically granted to students with special merits. In the 2004 VHLSS the total estimated value of scholarships is only 565 billion VND, so we do not include scholarships in the present analysis.

There are virtually no tuition fee exemptions at the primary school level. The reason is that tuition is now free of charge at the primary level in all public schools. Hence, most of the tuition fee exemption expenditures benefit students in lower and upper secondary schools, with slightly higher total benefits at the lower secondary level (46 and 38 percent of the total value of fee exemptions, respectively). In contrast, for contribution fee exemptions the largest share benefits primary level students (43 percent) followed by lower secondary students (33 percent) and a much smaller part to upper secondary students (17 percent). Consequently, based on aggregate numbers, education fee exemptions may well be pro-poor as large shares of the benefits are targeted towards the 'right' groups: primary and lower secondary students in the rural area.

In the following sections we move beyond the totals to look at the incidence of the social transfers and education fee exemptions using the 2002 and 2004 household surveys. In the analysis of education fee exemptions we exclude tuition fee exemptions granted to primary level students as these exemptions are in all likelihood granted for quite special reasons given that primary level tuition is free of charge in public schools.

#### 3. The Behavioral Response to Social Transfers and Education Fee Exemptions

When looking into the incidence of the social transfers and education fee exemptions a key issue is the computation of a counterfactual welfare indicator. The problem is not so much in the definition of the welfare indicator as most studies use either total income or total consumption expenditure. The problem is rather how to define and compute the counterfactual. Most conventional benefit incidence analyses are "non-behavioral" in the sense that they assume there is no substitution of any kind in response to a transfer (see e.g., Demery, 2003). Hence, in the absence of transfers households would simply consume an amount equal to the observed total expenditure less the monetary value of the transfer. Whether such a 'pre-intervention' expenditure measure is an appropriate counterfactual to take for assessing the incidence of public spending is a recurrent theme in incidence analyses. For the assumption to be acceptable, the observed distribution of income and expenditure must be unaffected by the public program that is analyzed. This is rarely the case. Yet, the public transfers analyzed in this chapter are relatively small in the sense that the total transfers constitute a small share the total income generation. Hence, relative prices are probably not sensitive to changes in social transfers and education fee exemptions.

Even in the absence of price effects there may be behavioral responses to social transfers at the household level. In particular, a transfer may increase savings rather than expenditures; it may reduce the labor supply; or it may affect private transfers such as remittances and private borrowing and lending. Ravallion, van de Walle and Gautam (1995) suggests a simple way of estimating the behavioral response to transfers at the household level by estimation of the marginal propensity to consume out of transfers in a reduced form expenditure regression.<sup>84</sup> van de Walle (2004a) uses that approach in her analysis of Vietnam's public safety net in the 1990s. In this study we also follow the suggestion and estimate the marginal propensity of consume out of social transfers (*mpct*) and the marginal propensity to consume out of education fee exemptions (*mpce*) using household level panel data.

#### 3.1. The Model and Results for Social Transfers

Given repeated observations for a set of households, a reduced form consumption model can be formulated as

<sup>&</sup>lt;sup>84</sup> See also van de Walle (2003) for a succinct description of the method.

$$E_{it} = T_{it}\beta + X_{it}\gamma + \alpha_i + \lambda_t + \varepsilon_{it},$$
<sup>(1)</sup>

where  $E_{it}$  is real expenditure in household *i* at time *t*,  $T_{it}$  is the transfer to household *i* at time *t*, and  $X_{it}$  are observed household characteristics. The unobservable characteristics are split into a time invariant household specific latent factor  $\alpha_i$ ; a common time varying latent factor  $\lambda_i$ , and an idiosyncratic term  $\varepsilon_{it}$ , as in most panel data analyses.

The parameter of interest,  $\beta$ , is a direct measure of the marginal propensity to consume out of transfer income, which is assumed to be the same for all households in equation (1). Estimation of this parameter is complicated by possible endogeneity problems that have to be dealt with. Endogeneity of transfers may arise because of targeting of transfers towards certain households with chronic problems that may lead to persistently low expenditures (e.g., households with invalids or gravely disabled members). Such targeting implies a negative correlation between transfers and the latent household specific factor ( $cov(T_{it}, \alpha_i) < 0$ ), which will cause a downward bias in the least squares estimator. A second problem is that some of the social protection transfers are designed to be responses to shocks (natural disasters). Such transfers will be correlated with the idiosyncratic components ( $cov(T_{it}, \varepsilon_{it}) < 0$ ) if the natural disaster has an impact on expenditures as one would expect for poor, credit constrained households. This correlation will also lead to a downward bias in the least squares squares estimator.

In the regressions reported below we deal with the endogeneity problems by transforming the regression model using changes from 2002 to 2004 (*i.e.*, the model is estimated in first differences). This transformation annihilates the household specific latent factors and hence the first cause of bias. Subsequently, we take account of the second endogeneity problem by applying an instrumental variable estimator.

The data are the panel component of the two surveys 2002 and 2004 VHLSS that includes 3,935 households. The dependent variable is real total household expenditure measured in January 2004 prices and the transfer variable is the sum of social insurance and social protection benefits both of which are measured at the household level.

The set of conditioning variables included in the regression models consists of basic measures of household composition and education. Information about land assets and occupation of household members could also have been included. However, it is important to keep in mind that by using changes from 2002 to 2004 all time invariant household attributes are implicitly included and annihilated by the difference transformation.<sup>85</sup> Summary statistics for the variables included in the regressions are given in Table A1 in Appendix 2.

Table 4 reports our basic regression results for the social transfers. In regression (1), changes in transfers is the only regressor apart from the intercept and, hence, this is the standard difference-indifference estimator. The estimated *mpct* is about one-third in regression (1). Next, in regression (2) the control variables are added to take account of observable time varying differences between households. In general, the signs and magnitudes of the control variables are sensible; household consumption expenditure is increasing in household size, but not linearly; the presence of children in a household has a negative effect on expenditure, given the household size, and consumption expenditure is increasing in the level of education of the household members. The most interesting result in regression (2) is that inclusion of the controls leads to a large drop in the *mpct* estimate, down to around one-fifth. Finally, when applying the GMM estimator in regression (3) the estimated *mpct* is 1.

<sup>&</sup>lt;sup>85</sup> Needless to say, this is also the case for other time invariant factors such as province, district and commune information.

Dependent variable: Real household expenditure, 1.000 VND January 2004 prices.						
	Social insurance				Only	
	an	and social protection			Social insura	nce
Regression	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	LS	LS	GMM	LS	LS	GMM
Transfer	0.328	0.197	0.988	0.315	0.196	1.028
	(3.29)	(2.49)	(2.14)	(3.18)	(2.48)	(2.20)
Size of household		3,388	3,580		3,395	3,623
		(6.33)	(6.72)		(6.34)	(6.74)
Size of household squared		-78	-81		-78	-84
·		(1.75)	(1.83)		(1.76)	(1.90)
Gender of household head		1,080	1,017		1,09Ź	1.091
		(1.36)	(1.29)		(1.38)	(1.37)
Household members who are		()	( - )		( )	( - )
Children 0-5		-1,059	-1,229		-1,060	-1,241
		(2.20)	(2.62)		(2.21)	(2.64)
Children 6-10		-782	-818		-787	-839
		(2.28)	(2.54)		(2.29)	(2.60)
Primarv-educated		-144	-208		-144	-198
,		(0.54)	(0.82)		(0.54)	(0.78)
Secondarv-educated		389	336		387	329
·····		(1.45)	(1.27)		(1.44)	(1.24)
High-school-educated		1.453	1.058		1.45Ó	1.035
3		(3.41)	(2.43)		(3.40)	(2.36)
Vocationally educated		2.825	2.334		2.826	2.337
·····		(4.85)	(4.09)		(4.85)	(4.11)
Professionally educated		3.455	3.217		3.454	3.194 <sup>́</sup>
,		(5.25)	(4.79)		(5.24)	(4.75)
University-educated		3.744	3.270		3.758	3.337
		(4.09)	(3.33)		(4.12)	(3.41)
Constant	3.231	3.033	2.920	3.240	3.037	2.924
	(15.67)	(15.48)	(13.87)	(15.71)	(15.49)	(13.96)
Observations	3.935	3.935	3.935	3.935	3.935	3.935
RESET test (p-value)	0.72	0.43	0.71	0.75	0.44	0.79
First stage <i>F</i> -statistic <sup>a</sup>			50.28			46.96
Hansen J-test <sup>b</sup> (p-value)			0.43			0.48

Table 4: The Estimated Marginal Propensity to Consume out of Social Transfers

*Notes*: The dependent variable and all regressors are given as changes from 2002 to 2004. All regressions use sample weights from the 2002 VHLSS survey. Cluster robust *t*-statistics in parentheses. Instruments used in regressions (3) and (6): Change in number of females above 55; change in number of males above 60; initial number of females above 55; and initial number of males above 60. <sup>a</sup>The first stage *F*-statistic is the Cragg-Donald test statistic for instrument relevance. <sup>b</sup>The Hansen *J*-test is the over-identifying restriction test. *Source:* Own calculations based on the 2002 and 2004 VHLSS.

In the GMM regression we use four instrumental variables; the changes in the number of males above 60 and females above 55 years of age from 2002 to 2004 (*i.e.*, the change in the number of pension age members in the household) and the initial number of males over 60 and females over 55 years of age (*i.e.*, the number of pension age members in the household in 2002). The choice of pension age household members as instruments is guided by two observations. First of all the change in the number of pension age members is in most cases predictable at the household level. Entering the pension age is clearly predictable and leaving pension age is often predictable within a certain time interval even if the precise dating is unknown. The predictability of the change means that households should smooth consumption across this event whereby it is not a shock to household expenditure.

This makes the variables valid instruments. Second, movements into and out of the pension age is associated with changes in pension transfers for individuals covered by the social insurance program as explained in Section 2. Hence, there is reason to assume that changes in pension age individuals are correlated with changes in social insurance transfers, at least for some households, making the variables relevant instruments. The first stage *F*-statistic reported for regression (3) in Table 4 confirms the assumption of relevant instruments while the test of over-identifying restrictions, the *J*-test, does not reject the validity of the instruments.

van de Walle (2004a) reports similar regressions based on the panel data component in the two household surveys 1993 and 1998 VLSS. However, she finds much smaller discrepancies between the three estimators and concludes that an *mpct* around 0.5 is reasonable as it is within the 95 percent confidence bound in all three regressions. This is in fairly sharp contrast to our results, as they seem to imply an *mpct* quite close to one.

There are some differences in the choice of conditioning variables between van de Walle (2004a) and our model. However, the main dissimilarities are in the use of GMM rather than TSLS and, in particular, in the choice of instruments in the IV-regression. Specifically, van de Walle uses the initial value of transfers as the sole instrument for the changes in transfers while we use the four instruments discussed above. We have also estimated models in which the initial value of transfers is included as an instrument. The results show that this instrument is irrelevant in our model formulation. This is in all likelihood caused by the large changes in the Vietnamese economy and changes in the administration of transfers from the mid 1990s to the early/mid 2000s.

In order to look more into the robustness of our results we re-estimate the three regression models using only social insurance benefits as the transfer variable.<sup>86</sup> As shown in Section 2, the social insurance benefits are by far the largest transfers and changes in social insurance transfers are more closely associated with changes in the number of pension age household members than the total transfers used in regressions (1)-(3) in Table 4. The results, when using only the social insurance transfers, are given as regressions (4)-(6) in Table 4. The latter estimates are very similar to the estimates in regressions (1)-(3) in which all transfers are included.

<sup>&</sup>lt;sup>86</sup> While 912 households in our sample received at least one of the transfers only 539 households received social insurance transfers.

Dependent variable: Real household expenditure, 1.000 VND January 2004 prices.								
	Social insurance				Only			
	and social protection			S	ocial insuran	се		
Regression	(1)	(2)	(3)	(4)	(5)	(6)		
Estimator	LS	LS	GMM	LS	LS	GMM		
Transfer	0.397	0.207	1.038	0.384	0.212	1.103		
	(2.62)	(1.74)	(1.99)	(2.49)	(1.74)	(2.04)		
Size of household		3,408	3,575	3,236	3,038	2,980		
		(6.37)	(6.73)		(6.38)	(6.75)		
Size of household squared		-79	-80		-80	-83		
		(1.78)	(1.80)		(1.79)	(1.88)		
Gender of household head		1,148	1,096		1,160	1,173		
		(1.44)	(1.39)		(1.45)	(1.47)		
Household members who are								
Children 0-5		-1,077	-1,255		-1,078	-1,274		
		(2.24)	(2.64)		(2.24)	(2.67)		
Children 6-10		-786	-821		-790	-840		
		(2.30)	(2.55)		(2.31)	(2.61)		
Primary-educated		-142	-193		-143	-192		
		(0.54)	(0.76)		(0.54)	(0.76)		
Secondary-educated		379	320		376	302		
		(1.41)	(1.21)		(1.40)	(1.13)		
High-school-educated		1,472	1,047		1,470	1,023		
		(3.40)	(2.33)		(3.39)	(2.25)		
Vocationally educated		2,810	2,355		2,794	2,340		
		(4.79)	(4.15)		(4.77)	(4.13)		
Professionally educated		3,515	3,141		3,515	3,122		
		(5.29)	(4.39)		(5.29)	(4.36)		
University-educated		3,738	3,476		3,746	3,519		
		(3.99)	(3.57)		(4.01)	(3.61)		
Constant	3,227	3,035	2,974	3,236	3,038	2,980		
	(15.65)	(15.52)	(14.69)	(15.69)	(15.52)	(14.79)		
Observations <sup>a</sup>	3,923	3,923	3,923	3,925	3,925	3,925		
RESET test (p-value)	0.43	0.41	0.59	0.39	0.41	0.65		
First stage F-statistic			68.6			63.07		
Hansen J-test (p-value)			0.34			0.37		

### Table 5: The Estimated Marginal Propensity to Consume out of Social Transfers: Robustness to Outliers

Notes: See the notes to Table 2. <sup>a</sup>Outliers in the transfer variables have been identified using Hadi's (1992,1994) iterative outlier detection procedure and these regressor outliers are omitted from the regressions.

Source: Own calculations based on the 2002 and 2004 VHLSS.

Finally, because of the aforementioned extreme transfer values in the surveys we identify outlying transfer observations, again using the Hadi algorithm, as explained in Section 2. Subsequently, all models in Table 4 are re-estimated based on a sub-sample that excludes the outliers. The results of the re-estimations, given in Table 5, indicate that our basic results are not driven by the extreme transfer observations. Hence, both robustness checks support our basic result.

#### 3.2. The Model and Results for Education Fee Exemptions

In the estimation of the marginal propensity to consume out of education fee exemptions (*mpce*) we use the same models and the same set of estimators as for social transfers. However, there is an important difference in the treatment of social transfers and education fee exemptions in the household surveys. All education fees and contributions paid by the household are included in total consumption expenditure (*E*). But when a household is granted exemption from a fee this expenditure is not included in total expenditure even though the student is consuming the education good. This means that the 'true' measure of household consumption ( $E^*$ ) is the sum of observed consumption expenditure and the value of the education fee exemption:

$$E_{it}^{*} = E_{it} + S_{it},$$
 (2)

where *S*<sub>*it*</sub> is the value of the education fee exemption. As our regression model for the true household expenditure is

$$E_{it}^* = S_{it}\beta + X_{it}\gamma + \alpha_i + \lambda_t + \varepsilon_{it}, \tag{3}$$

we can use equation (2) to substitute for the true household expenditure, resulting in the model

$$E_{ii} = S_{ii}(\beta - 1) + X_{ii}\gamma + \alpha_i + \lambda_i + \varepsilon_{ii}.$$
(4)

Hence, in the education fee exemption regressions a parameter estimate of 0 corresponds to an *mpce* of 1 and an estimate of, say, -0.5 corresponds to an *mpce* of 0.5.<sup>87</sup>

Table 6 reports our results for education fee exemptions. The fee exemption regressor is the sum of the monetary values of tuition fee exemption and contribution fee exemption. Both fee exemptions are measured at the household level in January 2004 prices. Regressions (1) to (3) in Table 6 corresponds to the first three regressions in Table 4. In all three regressions we find an estimated coefficient of zero, indicating that the *mpce* is one.

<sup>&</sup>lt;sup>87</sup> Another way to think of this result is to consider a household that is granted tuition fee exemption. If the consumption of that household is unchanged (*mpce* = 0) then the observed household expenditure will drop by the value of the fee exemption. If the households chooses to spend exactly what is saved on tuition (*mpce* = 1) then the observed household expenditure will return to the pre tuition fee exemption level.

Dependent variable: Real household expenditure, 1.000 VND January 2004 prices.						
· · ·		All observati	ons	. (	Dmitting out	iers
Regression	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	ĹŚ	ĹŚ	GMM	ĹŚ	LŚ	GMM
Education reduction/exemption	0.005	-0.019	0.032	0.062	0.014	0.033
	(0.15)	(0.61)	(1.37)	(2.43)	(0.65)	(1.44)
Size of household		3,377	3,360		3,391	3,389
		(6.28)	(6.27)		(6.27)	(6.27)
Size of household squared		-76	-75		-79	-79
		(1.69)	(1.68)		(1.73)	(1.74)
Gender of household head		1,099	1,113		1,118	1,118
		(1.39)	(1.41)		(1.41)	(1.41)
Household members who are						
Children 0-5		-1,056	-1,046		-1,062	-1,053
		(2.19)	(2.18)		(2.20)	(2.19)
Children 6-10		-790	-788		-735	-732
		(2.29)	(2.28)		(2.16)	(2.15)
Primary-educated		-150	-151		-126	-125
		(0.56)	(0.56)		(0.47)	(0.47)
Secondary-educated		388	390		389	387
		(1.44)	(1.45)		(1.45)	(1.44)
High-school-educated		1,505	1,516		1,474	1,480
		(3.56)	(3.58)		(3.47)	(3.49)
Vocationally educated		2,889	2,887		2,854	2,859
		(4.96)	(4.98)		(4.88)	(4.90)
Professionally educated		3,566	3,486		3,499	3,482
		(5.22)	(5.05)		(5.10)	(5.05)
University-educated		3,824	3,855		3,888	3,901
		(4.36)	(4.35)		(4.40)	(4.41)
Constant	3,318	3,083	3,070	3333	3,103	3,099
	(15.94)	(15.55)	(15.78)	(16.06)	(15.78)	(16.13)
Observations	3,935	3,935	3,935	3,908	3,908	3,908
RESET test (p-value)	0.11	0.43	0.42	0.37	0.43	0.43
First stage F-statistic			739.6			3,952
Hansen J-test (p-value)			0.90			0.90

### Table 6: The Estimated Marginal Propensity to Consume out of Education Fee Exemption

*Notes*: The dependent variable and all regressors are given as changes from 2002 to 2004. All regressions use sample weights from the 2002 VHLSS survey. Cluster robust *t*-statistics in parentheses. Instruments used in regressions (3) and (6): Initial tuition fee exemptions and initial school contribution fee exemptions. See the notes to Tables 2 and 3 for explanations of the test statistics and the outlier detection.

Source: Own calculations based on the 2002 and 2004 VHLSS.

In the GMM regression we cannot use the most relevant household composition variables as in the regressions for social transfers because children have a direct influence on consumption expenditure. Instead, we use the initial levels of the tuition fee exemption and the contribution fee exemption as instruments. This implies an assumption of predetermined education fee exemptions (  $E(\varepsilon_{it} | S_{it}) = 0$ ), which we do not find unreasonable even though we would have preferred a more structural approach to the choice of instruments. Judged by the *F*-test for instrument relevance and the *J*-test for overidentifying restrictions the chosen instruments are (highly) relevant and not invalid.

Table A1 in Appendix 2 reveals that some of the changes in education fee exemptions are even more out of the ordinary than the extreme changes in social transfers. Therefore, we, again, identify and exclude outlying observations using the Hadi algorithm and re-run the estimations. The results of the re-estimations are reported as regressions (4)-(6) in Table 6. Exclusion of the outliers does not change the basic result of an *mpce* of one.

In sum, our regressions point to a marginal propensity to consume out of both social transfers and education fee exemptions of one. However, the *mpct* is not precisely determined, so even though the point estimate is one, we cannot reject the hypothesis that the true value is only one-half as found in van de Walle (2004a). In fact, based on the GMM results in Tables 2 and 3 the *mpct* can be anywhere from about 0.1 to more than 1. This means that the social transfer incidence results presented in the subsequent sections must be interpreted with great caution. In contrast, all regressions using education fee exemptions in Table 6 firmly rejects an *mpce* below 0.9. Therefore, we have more confidence in using an *mpce* of one in the incidence analysis of the education fee exemptions.

#### 4. The Incidence of Social Transfers and Education Fee Exemptions in 2004

In the incidence analysis, we use the information from the 2004 VHLSS about the transfers and education fee exemptions that households received during the last twelve months since the date of interview. Based on this information we calculate per capita transfers and per capita education fee exemptions as the household benefit divided by the household size. This implies that all members of a household will benefit from, say, pensions, if one member receives that kind of transfer and, likewise, all members benefit from, say, tuition fee exemptions, if such exemptions are granted to a student the household. These per capita measures correspond to the computation of real per capita consumption expenditure in the survey.

Our basic measure of welfare is total per capita consumption expenditure as it is measured in the 2004 VHLSS. In the incidence analysis we compute two counterfactual welfare measures, one for social transfers and one for education fee exemptions, based on the results in section 3. Hence, for social transfers we use per capita consumption expenditure net of per capita social transfers as the preferred counterfactual welfare measure (*i.e.*, an *mpct* of one). For education fee exemption we use the actual per capita consumption expenditure, which is implying an *mpce* of one. However, we also look into the robustness of our results by assuming an *mpct* of 0.5. Using the lower marginal propensity to consume out of transfers, we obtain results that are directly comparable with the analyses reported in van de Walle (2004a,b), the World Bank (2002) and Justino (2005).

In the incidence tables the population is divided into quintiles based on the counterfactual per capita consumption expenditure and the within quintile average transfer/fee exemption is given. In addition, we also decompose the average transfer/fee exemption in order to obtain a more detailed analysis of the social transfer and education fee exemption programs. Specifically, we use the decomposition

$$\overline{T}_q = \frac{1}{N_q} \sum_{i=1}^{N_q} T_{qi} = \left(\frac{R_q}{N_q}\right) \left(\frac{B_q}{R_q}\right) \left(\frac{1}{B_q} \sum_{i=1}^{B_q} T_{qi}\right), \quad q = 1, \dots, 5,$$
(5)

where  $\overline{T}_q$  is the average transfer/fee exemption in quintile q,  $N_q$  is the number of individuals in the quintile,  $R_q$  is the number of individuals in the quintile who are in the reference (or target) population, and  $B_j$  is the number of individuals the quintile who actually receive a transfer or are granted education fee exemption (the beneficiaries). Finally, the summation in the last factor is only over the

non-zero social transfers and education fee exemptions, so this is the average transfer/fee exemption to the beneficiaries.<sup>88</sup>

Looking across the quintiles, the first factor on the right-hand side of equation (5),  $(R_q/N_q)$ , is the distribution of the reference population within the total population. That distribution shows to what extent the social transfer and education fee exemption programs are "designed" to be targeted towards the poorer individuals from the outset. In the definition of the reference populations we do not include information about the level of per capita consumption. Instead, we aim at capturing broader criteria. In particular, we mainly use information about the age of household members even though more precise criteria could be used.

The second factor is the coverage rate for each program  $(B_q/R_q)$ . The distribution across quintiles illustrates if the coverage rate varies with the expenditure level, which may indicate good or bad targeting in the administration of the programs. The product of the reference population rate and the coverage rate is the conventional incidence  $(B_q/N_q)$  of the program.

Finally, as the third factor is the average transfer or education fee exemption for the beneficiaries the distribution across quintiles shows if the poorer beneficiaries get higher transfers than the richer beneficiaries or vice versa.

The decomposition in equation (5) also applies when transfers and education fee exemptions as expressed as shares of per capita consumption expenditures. Therefore, in the tables below, we only report the average of the share of transfers and education fee exemptions in total consumption expenditure for the beneficiaries. When computing these shares we use the actual per capita consumption expenditure throughout.

In addition to the decomposition of the average transfer we also report results for the two subpopulations "Rural Vietnam" and "Urban Vietnam". The reason for our interest in this split is given from Table 1 and Table 3, in which one finds a clear urban bias in the largest program; social insurance, and a (smaller) rural bias in the three programs; social protection, tuition fee exemption and contribution fee exemption.

Finally, we stress, again, that all results are based on data in which we have removed extreme transfer and education fee exemption observations using the Hadi (1992, 1994) algorithm.

#### 4.1. Social Transfers

The incidence of social insurance transfers is shown in Table 7. The distribution of social insurance is 'U-shaped' in the sense that the poorest and the richest quintiles are the prime beneficiaries of the program in terms of average transfers. The average transfer to the poorest quintile is about 350 thousand VND per capita per year, while it is about 230 thousand VND to the richest quintile. Average transfers to the three middle quintiles are just above one-third of the average transfer to the poorest quintile. Using the conventional measure of redistribution, one finds that the social insurance transfers are progressive as the share of transfers in total expenditure is highest in the poorest quintile (10 percent). The share of transfers in total expenditure drops to around 2.5 percent in the second and third quintiles and just below 2 percent in the two richest quintiles.

<sup>&</sup>lt;sup>88</sup> The decomposition can be given generally as a product of conditional expectations,  $E(T_q|C_q)=E(T_q|B_q,R_q,C_q)E(B_q|R_q,C_q)E(R_q|C_q)$ . Hence, each factor in (5) has a well defined population mean.

	Quintiles					
	(Poorest)		(Richest)			
	Q1	Q2	Q3	Q4	Q5	All
All Vietnam			Mpct = 1			
Average transfer	349.39	117.03	135.64	127.88	228.34	191.66
Reference population	41.21	32.53	29.82	32.45	31.37	33.47
Coverage rate	39.44	21.47	28.14	24.49	37.83	30.73
Average transfer to recipients	2149.77	1675.96	1616.67	1608.8	1924.23	1862.88
Average share of expenditure	61.93	35.13	29.53	23.09	15.89	36.39
Rural Vietnam						
Average transfer	327.77	77.10	78.20	73.22	86.24	128.54
Reference population	41.52	32.77	29.15	29.20	29.91	32.51
Coverage rate	37.00	20.16	21.38	19.26	21.64	24.79
Average transfer to recipients	2133.89	1166.93	1254.78	1302.37	1332.55	1594.77
Average share of expenditure	64.70	31.14	26.97	24.37	16.64	40.02
Urban Vietnam						
Average transfer	669.02	327.46	263.87	251.59	330.53	368.59
Reference population	42.84	38.28	32.99	36.36	28.18	35.73
Coverage rate	57.99	39.67	40.92	34.68	54.43	45.61
Average transfer to recipients	2693.28	2156.06	1954.94	1995.08	2154.46	2261.49
Average share of expenditure	56.75	30.60	22.39	17.65	12.29	31.77
All Vietnam			Mpct = 0.3	5		
Average transfer	153.55	135.94	136.08	205.88	326.85	191.66
Reference population	38.37	32.61	30.13	33.99	32.27	33.47
Coverage rate	23.83	24.03	27.75	32.64	46.49	30.73
Average transfer to recipients	1679.03	1734.51	1627.18	1856.02	2178.51	1862.88
Average share of expenditure	63	45.4	35.43	30.96	20	36.39
Rural Vietnam						
Average transfer	136.47	109.36	101.97	145.86	149.03	128.54
Reference population	39.50	31.49	30.10	30.39	31.07	32.51
Coverage rate	21.87	22.90	21.42	28.39	30.18	24.79
Average transfer to recipients	1580.11	1516.17	1581.44	1690.64	1589.70	1594.77
Average share of expenditure	63.80	44.67	38.08	33.15	22.19	40.02
Urban Vietnam						
Average transfer	371.81	314.13	334.13	384.08	439.24	368.59
Reference population	38.82	38.70	33.58	38.06	29.54	35.73
Coverage rate	41.44	39.83	44.85	44.75	60.67	45.61
Average transfer to recipients	2311.29	2038.32	2218.81	2255.37	2450.46	2261.49
Average share of expenditure	60.15	35.74	28.20	22.26	14.84	31.77

**Table 7: Incidence of Social Insurance** 

*Notes:* The quintiles are based on actual per capita consumption expenditure less per capita social insurance and per capita social protection transfers. All monetary values are given in 1.000 VND, January 2004 prices. Rates and shares are given as percentages. *Source:* Own calculations based on 2004 VHLSS.

In the decomposition of the average transfers we use the population of all pension-age individuals as the reference population (*i.e.*, females above 55 years of age and males above 60). This is clearly not a precise approximation as only a small fraction of the pension-age population is covered by the social insurance system. However, the 2004 VHLSS does not include information about pre-retirement occupation, so we cannot distinguish social insurance members from non-members in the survey. Another, problem with the definition of the reference group is that short term support for sickness

and maternity is given to individuals in the working age population who are covered by the program. We find this problem to be of minor importance as the total benefits given for short term support are small relative to the pension transfers. Moreover, the distribution of the pension-age population across the welfare distribution is interesting in its own right as it illustrates if a broader (unitary) pension system will be pro-poor. Finally, the incidence, calculated as the product of the reference population and the coverage rate is not affected by the error in the definition of the reference population.

The distribution of the pension-age population across the expenditure quintiles demonstrates that a general unitary pension scheme will only be mildly pro-poor. About 40 percent of the individuals in the poorest quintile live in households with a pension-age member. In the four richer quintiles the share is close to 30 percent with no systematic tendency to either increase or decrease across quintiles. Thus, while the poor may often be old-age people one cannot conclude that old-age people are often poor, as it is sometimes done in other studies.

The coverage rate is the prime explanation for the U-shaped distribution of the social insurance transfers. The coverage rate is highest in the poorest quintile (40 percent) followed by the richest quintile (38 percent), and somewhat lower rates for the three middle quintiles (21-28 percent). Combining the reference population and the coverage rate we find a U-shaped distribution of the incidence as the rate is 16 percent in the poorest quintile, 12 percent in the richest quintile and about 8 percent in the three middle quintiles.

Another cause of the U-shaped distribution can be found in the distribution of the size of the transfers to the beneficiaries. The individuals in the poorest quintile benefit most while the richest quintile comes in second. On average, the recipients in the lowest quintile receive slightly more than 2.1 million VDN per capita per year. This is as surprisingly large amount considering the official GSO poverty line of 2.077 million VND (January 2004 prices), and it implies that the average recipient in the lowest quintile is non-poor—after the transfer—given our estimate of the marginal propensity to consume out of transfers.<sup>89</sup> The size and importance of the transfer is underlined by the share of the transfer in total expenditure. In the poorest quintile the share is almost two-thirds; in the second quintile (receiving somewhat less) the share is more than one-third and even in the fourth quintile the share is close to one-quarter.

The decomposition of the individuals into Rural and Urban Vietnam does not change the over-all picture of redistribution. However, the urban bias in the system is clear when looking both at the coverage rates and the average transfer to the recipients. While the coverage rates for Rural Vietnam are close to the overall averages in the four lowest quintiles and much smaller for the richest quintile (22 percent compared to 38 percent), the coverage rates in Urban Vietnam are much higher. In the poorest and richest quintiles in Urban Vietnam, the coverage rates are above 50 percent, leading to incidence rates of 25 percent in the poorest quintile and 15 percent in the richest quintile. Moreover, average transfers to the recipients are much higher in Urban Vietnam compared to Rural Vietnam. The latter result is not surprising considering the urban bias in the selection into the security system and the proportionality of benefits and past wages.

Given the large share of state funding in the social security system and the relatively small amount of resources allocated to the social protection system and the national targeted programs, one may ask if the present social insurance system is too generous. Nevertheless, the system does reach the poorest households and the system has a clearly progressive redistributive profile, given the choice of counterfactual per capita consumption expenditure.

In the bottom part of Table 7 the distribution of social insurance transfers is repeated using the alternative welfare indicator that is preferred by van de Walle (2004a, b), the World Bank (2003) and

<sup>&</sup>lt;sup>89</sup> When looking at the average transfers to recipients it is important to remember that a household may have more than one beneficiary. As already mentioned, the numbers we report are the average per capita transfers based on information about total social insurance transfers to households.

Justino (2005). The alternative welfare indicator results in a much less progressive distribution of the social insurance transfers. This is not surprising considering the size of the transfers to the recipients. In particular, the change causes a shift in the transfers away from the bottom quintile and towards the two top quintiles both in terms of coverage rate, transfer to recipients and, hence, also the overall average transfer to the quintile populations.

Turning next to the social protection system, we report the results of the incidence analysis in Table 8. As expected, the sizes of the average social protection transfers are of a much smaller magnitude than the social insurance transfers. The social protection transfers are clearly progressive, but not as much as one would expect given the design of the regular relief program. In particular, the average transfer to the poorest quintile is only five times the average transfer to the richest quintile.

			Quintile	S		
	(Poorest)				(Richest)	
	Q1	Q2	Q3	Q4	Q5	All
All Vietnam			mpct = 1			
Average transfer	49.11	31.3	20.04	24.4	10.69	27.11
Reference population	100	100	100	100	100	100
Coverage rate	16.88	8.63	5.14	5.81	2.90	7.87
Average transfer to recipients	290.93	362.79	390.29	420.06	368.76	344.44
Average share of expenditure	13.9	11.62	9.84	7.97	4.07	11.27
Rural Vietnam						
Average transfer	54.40	31.02	25.58	20.09	22.74	30.77
Reference population	100	100	100	100	100	100
Coverage rate	19.26	10.10	6.10	5.13	5.30	9.18
Average transfer to recipients	282.45	307.18	419.45	391.29	428.89	335.17
Average share of expenditure	14.02	10.83	11.97	9.10	6.87	11.67
Urban Vietnam						
Average transfer	26.83	29.05	6.51	9.36	6.71	15.70
Reference population	100	100	100	100	100	100
Coverage rate	9.65	7.19	2.21	3.24	1.56	4.77
Average transfer to recipients	278.02	404.07	293.97	288.68	431.20	329.00
Average share of expenditure	9.85	7.69	4.40	3.21	2.77	7.33
All Vietnam			mpct = 0	.5		
Average transfer	35.63	29.89	25.79	32.5	11.74	27.11
Reference population	100	100	100	100	100	100
Coverage rate	15.41	8.16	5.94	6.79	3.05	7.87
Average transfer to recipients	231.22	366.56	433.89	478.3	384.95	344.44
Average share of expenditure	12.71	12.81	11.56	9.03	4.32	11.27
Rural Vietnam						
Average transfer	37.19	29.04	27.56	31.68	28.38	30.77
Reference population	100	100	100	100	100	100
Coverage rate	17.30	9.93	6.02	6.77	5.86	9.18
Average transfer to recipients	214.92	292.33	457.67	467.60	483.97	335.17
Average share of expenditure	12.56	11.17	14.19	11.05	7.99	11.67
Urban Vietnam						
Average transfer	21.96	24.82	14.22	9.24	8.24	15.70
Reterence population	100	100	100	100	100	100
Coverage rate	9.07	5.97	3.28	3.88	1.67	4.77
Average transfer to recipients	242.11	415.84	433.33	238.25	494.30	329.00
Average share of expenditure	9 72	8 22	6 61	271	3 21	7 33

#### **Table 8: Incidence of Social Protection**

*Notes:* The quintiles are based on actual per capita consumption expenditure less per capita social insurance and per capita social protection transfers. All monetary values are given in 1.000 VND, January 2004 prices. Rates and shares are given as percentages. *Source:* Own calculations based on the 2004 VHLSS.

As explained in Section 2, the social protection program includes support to disaster-struck areas and, in theory, this could be anywhere in Vietnam. This is the reason why we have chosen to let the reference population be the whole population such that we simply report the reference population fraction as 100 percent. This means that the coverage rate is the incidence of the social protection transfers. The social protection program supports 17 percent of the population in the poorest quintile. This is not a large fraction considering that the GSO poverty rate for 2004 is just below 20 percent. However, it is important to recall the limited resources for the social protection program and that the regular relief transfers are only given to the extremely poor.

The most problematic result in Table 8 is the distribution of transfers to the recipients, which is increasing from the poorest to the richest quintile. Even if the transfers to the more well-off households are caused by natural disaster shocks it is hard to understand why the transfers to these households should be larger than transfers to the poorer households in the area.

The split into Rural and Urban Vietnam confirms the rural bias in the program; the average rural household receives twice the amount transferred to the average urban household. This difference is explained by the average incidence which is twice the size in the rural area compared to the urban area, while the average transfers to the recipients are roughly the same. The incidence in the poorest quintile in Rural Vietnam is higher than for All Vietnam, but it is still only 20 percent; four times the incidence of the richest quintile in Rural Vietnam. In Urban Vietnam the targeting of the social protection program appears to be a little better as only small fractions of the individuals in the richer quintiles receive transfers. Still, in both sub-populations we find the tendency of a positive association between welfare and transfers for the beneficiaries.

Using the alternative welfare indicator (the bottom part of Table 8) causes changes in the incidence of social protection transfers that are analogues to the results for social insurance transfers. The targeting towards the poorest individuals is worse than in the base case and the overall program is thus less progressive. However, the results for social protection appear to be less sensitive to the choice of counterfactual welfare measure relative to the social insurance transfers.

There are two, potentially serious, problems in our, and previous, assessments of the social protection program. The first problem is that the extremely poor, *i.e.*, the lonely elderly, orphans below 16, and severe invalids who cannot live on their own, are not well covered in multi-purpose household surveys like the 2004 VHLSS. This problem may cause a negative bias in the estimated incidence in the lowest quintile. However, this does not explain the relatively high incidence and the high transfers to beneficiaries in the richer quintiles.

But this may be explained by the second problem; our inability to separate transfers from the Social Guarantee Fund for War Affected Groups and transfers from the two other social protection funds. Consider, for example, a Vietnam Mother Hero. She receives 8.5 million VND a year from the state. In a household of 4, this is a transfer of 2.125 million VND per person per year, which is independent of household income. As we have no information about the relative size of the total transfers from the three social protection funds, we cannot exclude the possibility that the distribution of social protection transfers in Table 8 is to a large extent showing the distribution of war heroes and relatives of people with Revolutionary merit. The simplest, and best, way to overcome the identification problem is to have publicly available statistics for transfers from each of the social protection funds and, possibly, to include questions about the cause for social transfers in future household surveys, much in the same way as it is done for education fee exemption.

#### 4.2 Education Fee Exemptions

The results for tuition fee exemptions and contribution fee exemptions are given in Table 9 and Table 10, respectively. In both tables the quintiles are based on the distribution of actual per capita consumption expenditure, as explained above.

From Table 9 it is clear that the tuition fee exemption program provides fairly large support to all but the richest quintile. The average tuition fee exemption is largest for the middle quintile (50 thousand VND) followed by the poorest quintile. But the second and fourth quintiles benefit almost as much as the poorest quintiles.<sup>90</sup> The reason for the particular distributional shape of the tuition fee exemptions is an interesting combination of the reference population, the coverage rate and the size of the fee exemptions granted to the beneficiaries.

			Quintile	S		
	(Poorest)	(Poorest)			(Richest	)
	Q1	Q2	Q3	Q4	Q5	All
All Vietnam	All tuition fe	e exemption	ns			
Average exemption	40.73	35.21	51.31	36.22	16.36	35.97
Reference population	15.61	29.06	36.04	41.95	47.22	33.97
Coverage rate	33.42	13.15	12.74	5.88	2.55	10.19
Average exemption for recipients	780.49	921.57	1117.62	1468.88	1357.64	1039.36
Average share of expenditure	49.53	36.93	32.37	30.2	17.4	37.2
	00.00	05 10	00.10	40.00	04.05	05.01
Average exemption	30.09	35.13	33.10	49.99	24.05	35.81
	13.00	24.40	32.79	39.30	44.10	30.90
Coverage rate	37.54	15.31	11.59	11.20	4.21	12.31
Average exemption for recipients	/08.4	938.94	8/2.04	1128.23	1292.53	941.57
Urban Vietnam	48.50	43.31	29.94	30.89	23.44	37.23
Average exemption	56.64	53.01	35.95	12.07	14.19	34.38
Reference population	26.28	40.84	45.45	46.93	50.88	42.07
Coverage rate	18.34	8.65	5.55	1.42	0.71	5.66
Average exemption for recipients	1175.15	1499.66	1425.41	1806.73	3952.99	1443.45
Average share of expenditure	45.20	33.11	22.78	23.33	29.08	35.15
All Vietnam	Directly pov	erty related	tuition fee e	exemptions		
Average exemption	37.71	22.73	24.07	13.99	2.97	20.29
Reference population	15.26	28.48	35.9	41.79	47.42	33.77
Coverage rate	29.34	8.76	5.65	2.51	0.73	6.15
Average exemption for recipients	842.22	911.44	1186.57	1335.94	863.78	976.53
Average share of expenditure	53.83	37.05	34.1	28.63	12.54	42.05
Rural Vietnam						
Average exemption	35.35	27.05	17.48	23.13	6.41	21.89
Reference population	13.8	24.43	32.79	39.3	44.18	30.9
Coverage rate	34.67	11.32	5.42	5.34	1.62	7.86
Average exemption for recipients	739.07	978.49	983.64	1101.57	892.44	901.11
Average share of expenditure	50.58	45.76	34	29.57	17.01	41.44
Urban Vietnam						
Average exemption	39.79	30.42	8.21	0	0	15.69
Reference population	26.28	40.84	45.45	46.93	50.88	42.07
Coverage rate	11.75	3.76	1.53	0	0	2.53
Average exemption for recipients	1288.47	1982.5	1179.54	0	0	1474.17
Average share of expenditure	53.94	43.25	17.19	0	0	46.05

#### **Table 9: Incidence of Tuition Fee Exemptions**

*Notes:* The quintiles are based on actual per capita consumption expenditure. All monetary values are given in 1.000 VND, January 2004 prices. Rates and shares are given as percentages.

Source: Own calculations based on the 2004 VHLSS.

<sup>&</sup>lt;sup>90</sup> In terms of expenditure shares, the distribution of tuition fee exemptions is mildly progressive, though.

As we are only including tuition fee exemptions granted to students in lower secondary school and above, the reference population is defined as individuals in households in which at least one member is studying at the lover secondary level or above. The distribution of the reference population across per capita consumption groups confirms the frequently found pattern that child education is increasing with welfare. In the poorest quintile, only 16 percent of the individuals live in a household where at least one child is attending lower secondary school or above. This share jumps to 29 percent in the second quintile and then increases steadily to almost 50 percent in the richest quintile. Thus, as is well known, support programs for education above the primary level have an inherent tendency to be regressive.<sup>91</sup>

The coverage rate is working to equalize the skewed distribution of the reference group as 33 percent of the individuals in households with above-primary level students in the poorest quintiles benefit from tuition fee exemptions. The coverage rate is only 12-13 percent in the second and third quintiles and it drops to very low numbers in the top quintiles. The net result of the difference in the distributions of the reference group and the coverage rate is a fairly flat incidence distribution with about 5 percent of the individuals benefiting in the three poorest quintiles; dropping to 2.5 and 1.2 in the two richest quintiles. Hence, in that sense the targeting of the program is not too bad.

Next, the value of the tuition fee exemptions to the beneficiaries is clearly skewed towards the richer quintiles. The average tuition fee exemption is only 780 thousand VND in the poorest quintile, rising to almost the double (1,469 thousand VND) in the fourth quintile. The reason for the skew distribution of the fee exemption value to the beneficiaries is that students in rich households are in more expensive schools than students in poorer households, *i.e.*, school expenditures are in general rising with household income. Behrman and Knowles (1999) find a positive relationship between school expenditure and school quality in Vietnam based on data from 1996. If that result still holds in 2004, as one would expect, then school fee transfers to the beneficiaries would seem to be a better choice than school fee exemptions as it would be possible to put a cap on transfers, making the richer households pay for higher quality choices.

The split into Urban and Rural Vietnam shows an urban bias in the fee exemptions. This bias is caused by higher school participation rates in the urban areas and higher values of fee exemptions to beneficiaries. The bias is to some extent offset by the coverage rates. The average coverage rate is 12 percent in the rural areas, while it is only 6 percent in the urban areas. In particular, almost 38 percent of the reference population the poorest quintile in Rural Vietnam benefit from tuition fee exemptions. Yet, the reference population is only 14 percent of the population in that quintile, so the incidence is still low.

In the Vietnamese household surveys, beneficiaries of education fee exemption are asked about the reason for the fee exemption. This provides an opportunity to look closer at the incidence of directly poverty oriented education fee exemptions. Specifically, we have gathered four of the nine possible reasons into a group of poverty oriented fee exemptions. The four reasons are (i) "being a poor household", (ii) "having a difficult situation" (iii) "belonging to an ethnic minority group", and (iv) "living in remote, especially difficult area".<sup>92</sup> As seen, we include both direct targeting of the poor and categorical targeting in the definition of direct poverty oriented education fee exemption.

The bottom part of Table 9 reports the incidence results when only the direct poverty oriented tuition fee exemptions are included. First of all, we find that only about half (56 percent) of the value of tuition fee exemptions are granted for reasons of poverty. Second, the poverty related tuition fee exemption is much better targeted towards the poorer household than the total. The coverage rate in

<sup>&</sup>lt;sup>91</sup> Needless to say, in the present set-up we cannot evaluate the impact of tuition fee exemptions on school attendance for the poorest quintiles. If this effect is large, the relevant reference group should have been all children in the age group 11-18, and the relevant analysis would have been the increase in school attendance. We are, however, analyzing the incidence conditional on the impact in the present study.

<sup>&</sup>lt;sup>92</sup> The other possible reasons for education fee exemption are (v) "being a fallen combatant relative", (vi) being in a disabled, sick veteran policy household, (vii) being a primary student, (viii) no tuition required, and (ix) other reasons.

the poorest quintile is 29 percent, almost equal to the coverage rate for the total tuition fee exemptions. However, the coverage rates for the richer quintiles are significantly lower. The fairly small, but still measurable, coverage rates in the top quintiles are probably caused by the categorical targeting of the poor. This is confirmed by the rural/urban split which shows that no individuals in the top two quintiles in Urban Vietnam benefit from poverty related tuition fee exemptions. Moreover, in the rural areas the value of the tuition fee exemptions to the beneficiaries is fairly constant, with only a mild tendency to increase with income. Interestingly, the sample split also shows that the maximum value of the tuition fee exemption to the beneficiaries in the rural area (1,101 thousand VND to the fourth quintile) is lower than the value in all of the urban quintiles. Schooling above the primary level is more expensive in the urban areas and this is reflected in the fee exemptions.

	Quintiles					_		
	(Poorest)					-		
	Q1	Q2	Q3	Q4	(Richest) Q5	All		
All Vietnam	All contribut	All contribution fee exemptions						
Average exemption	100.47	59.07	46.97	17.88	15.38	47.96		
Reference population	81.71	80.53	79.51	75.71	75.54	78.60		
Coverage rate	18.75	9.52	6.39	2.81	1.67	8.00		
Average exemption for recipients	655.94	770.61	924.05	841.97	1222.83	762.43		
Average share of expenditure	45.33	31.89	27.14	18.06	14.1	36.03		
Rural Vietnam								
Average exemption	103.77	70.31	45.12	49.61	15.83	56.93		
Reference population	82.89	81.36	79.52	79.18	72.85	79.16		
Coverage rate	20.42	10.73	7.43	6.15	2.95	9.75		
Average exemption for recipients	613.15	805.07	763.77	1018.02	736.55	737.64		
Average share of expenditure	45.21	37.17	26.56	27.52	12.3	36.47		
Urban Vietnam								
Average exemption	62.43	17.9	23.58	4.77	17.72	25.29		
Reference population	74.66	75.88	80.8	77.07	76.64	77.01		
Coverage rate	11.75	2.82	2.37	0.46	0.66	3.56		
Average exemption for recipients	711.85	835.84	1229.26	1353.97	3502.27	923.13		
Average share of expenditure	33.86	17.79	19.24	16.01	24.24	28.49		
All Vietnam	Directly pov	erty related	l contribution	n fee exempti	ons			
Average exemption	92.29	47.06	26.78	8.75	5.44	36.07		
Reference population	81.94	80.23	79.48	75.74	75.64	78.61		
Coverage rate	16.77	7.14	3.38	1.45	0.61	6.04		
Average exemption for recipients	671.64	820.95	997.74	795.22	1182.27	760.25		
Average share of expenditure	47.48	33.53	29.96	18.02	14.84	40.13		
Rural Vietnam								
Average exemption	98.47	49.98	34.1	26.17	7.26	43.2		
Reference population	82.89	81.36	79.52	79.18	72.85	79.16		
Coverage rate	18.39	7.83	5.02	2.96	1.22	7.29		
Average exemption for recipients	646.12	784.41	853.59	1116.41	816.19	748.87		
Average share of expenditure	48.27	35.99	29.97	30.21	14.89	40.53		
Urban Vietnam								
Average exemption	61.16	4.04	7.48	0	4.41	15.43		
Reference population	74.66	75.88	80.8	77.07	76.64	77.01		
Coverage rate	9.91	0.74	0.97	0	0.19	2.31		
Average exemption for recipients	826.4	716.94	957.66	0	3000	866.83		
Average share of expenditure	40.77	16.07	13.84	0	14.94	36.41		

#### Table 10: Incidence of Contribution Fee Exemption

*Notes:* The quintiles are based on actual per capita consumption expenditure. All monetary values are given in 1.000 VND, January 2004 prices. Rates and shares are given as percentages.

Source: Own calculations based on the 2004 VHLSS.

The contribution fee exemptions, shown in Table 10, are also progressive, both as a share of consumption expenditure and in terms of average fee exemption values. The latter is in contrast to the three other types of support, in which there was no clear progression in the average values. In addition, there is a fairly clear poverty orientation as the average exemption value in the poorest quintile is almost twice the size of the value in the second quintile.

The decomposition of the average contribution fee exemptions shows that they need not be poverty oriented as the reference population is 75-82 percent in all quintiles, with the lower quintiles having only marginally higher shares. The reason for the relatively uniform distribution is that all students are included in the reference population, *i.e.*, we include primary level students as these are not automatically exempted from contribution fees. The coverage rates for contribution fee exemptions are much lower than tuition fee exemptions, but because of the larger reference population, the overall incidence is much higher in the poorest quintile (7 percent for contribution fee exemptions versus 2.6 percent for tuition fee exemptions). Next, the value of the fee exemption is increasing with per capita consumption expenditure. Hence, the poverty orientation is again caused by the coverage rate, and reduced by the value of the contribution fee exemption granted to the beneficiaries. Finally, also for contribution fee exemptions we find that it constitutes a large share of total per capita consumption expenditure for the poorer beneficiaries, and even for middle income beneficiaries (the third and fourth quintiles). So, education fee exemptions are important subsidies both to the poorer households in Vietnam and to some middle income households.

The rural/urban split is again interesting as it shows a rural bias, reflecting the total values in Table 3. This is caused both by higher reference population rates in the two poorest rural quintiles compared to the two poorest urban quintiles and by (much) higher coverage rates.<sup>93</sup>

Moving to the directly poverty oriented contribution fee exemptions in the bottom part of Table 10 we find that these fee exemptions take up 75 percent of the total value of contribution fee exemptions. This explains why the distribution of all contribution fee exemptions is more progressive than all tuition fee exemptions. It also explains the relatively minor differences between the distributions of directly poverty oriented contribution fee exemptions and all contribution fee exemptions. Yet, the coverage rates for the richest three quintiles decrease significantly (in relative terms), while the decrease in the coverage rates in the two poorest quintiles is relatively small, reflecting the poverty orientation. An 'oddity' in the results is the large fee exemption granted to 0.2 percent of the richest urban quintile, but this result does not qualitatively change the picture of fairly good targeting in the urban area, in which only the poorest quintile has a coverage rate above one percent.

Overall, both tuition fee exemptions and contribution fee exemptions are progressive, and fairly well targeted, given that the programs use both direct and categorical targeting. The division into poverty oriented fee exemptions and other fee exemptions shows that the poverty related exemptions are mainly benefiting the poorer households in terms of coverage. The main problem with the program is that the government cannot control the size of the subsidy, and there is a clear tendency for richer households to pay higher education fees, thereby benefiting from larger support when they are exempted from the fees. As already noted, one solution could be to replace the fee exemption system by a transfer system. This would make it possible to have a maximum subsidy controlled by the government. An additional benefit from an administrative change is that it would increase the transparency of the support program as the subsidies would be on the budget as expenditures, rather than in the current system in which the subsidies are revenues foregone.

<sup>&</sup>lt;sup>93</sup> There is a difference between these quintiles both in terms of consumption expenditure, as the poorest urban households are richer than the poorest rural households, and in terms of the number of children as poorer rural households, in general, have more children than poorer urban households.

#### 4.3. The Joint Incidence of Social Transfers and Education Fee Exemptions

We end this section by looking at the joint incidence of social transfers and education fee exemptions. In the combined analysis we use the counterfactual welfare measure from the social transfer analysis. That is, the welfare distribution is based on actual per capita consumption expenditure net of per capita transfers. Clearly, this will influence the incidence of education fee exemptions, but the choice of counterfactual is consistent with assuming that both of the marginal propensities to consume are equal to one.

As the sum of the transfers and education fee exemptions potentially targets all individuals in the population we do not look at the reference population or the coverage rate. Instead, we report the combination of the two, which is the incidence of the support. Moreover, we report the estimated total number of people benefiting from the government support systems.

Table 11 presents the incidence of the joint support system, which is clearly progressive when the subsidies are measured as shares of per capita consumption and in terms of actual support the poorest quintile receives most followed by the richest quintile. About 44 percent of the individuals in the poorest quintile are covered by some form of support and, on average, these beneficiaries receive 1.2 million VND, which is just around half of the poverty line in 2004. The incidence is 23 percent in the second quintile and falling slightly in the richer quintiles. The decline in the incidences is, again, counteracted by increasing subsidies to the beneficiaries from the second to the fifth quintile.

The estimates show that almost 24 percent of the whole population benefits from the joint government support. When the share is converted to actual numbers it shows that almost 19 million people are supported by either social transfers or education fee exemptions. Almost 40 percent of the people covered are in the poorest quintile and more than half (57 percent) are in the bottom 40 percent of the per capita consumption distribution.

	Quintiles					
	(Poorost) (Dichost)					
	(F001e31	, 	03	04	05	<i>)</i> ΔΙΙ
All Vietnam	Gi	QL.	QU	<b>U</b> 7	40	<u> </u>
Average subsidv	541.73	242.6	244.83	214.98	262.69	301.37
Incidence	43.79	23.19	19.06	16.47	15.03	23.51
Average subsidy for recipients	1237.25	1046.33	1284.6	1305.3	1747.38	1282.03
Average share of expenditure	49.71	32.45	29.47	22.24	15.24	34.77
Beneficiary population size	6.969	3.691	3.033	2.622	2.393	18.708
Rural Vietnam						
Average subsidy	531.5	216.42	186.41	180.67	145.02	252.05
Incidence	45.23	25.27	18.58	16.24	13.55	23.78
Average subsidy for recipients	1175.11	856.33	1003	1112.6	1070.31	1059.97
Average share of expenditure	50.66	32.58	28.11	25.77	15.25	35.86
Beneficiary population size	5.353	2.990	2.198	1.920	1.602	14.062
Urban Vietnam						
Average subsidy	813.64	444.37	312.19	279.79	369.09	443.97
Incidence	39.92	23.93	17.86	15.77	16.08	22.72
Average subsidy for recipients	2037.94	1857.31	1748.1	1773.7	2295.06	1954.07
Average share of expenditure	50.05	30.17	21.18	16.29	13.41	31.45
Beneficiary population size	1.634	0.981	0.730	0.643	0.658	4.647

Table 11: Incidence of Social Transfers and Education Fee Exemptions

*Notes:* The quintiles are based on actual per capita consumption expenditure less per capita social insurance and per capita social protection transfers. All monetary values are given in 1.000 VND, January 2004 prices. Rates and shares are given as percentages. The beneficiary population size is given in millions.

Source: Own calculations based on the 2004 VHLSS.

It is also interesting to look at the distribution of beneficiaries across Rural and Urban Vietnam. Some 75 percent of the beneficiaries (14 million people) live in the rural areas, so the joint system has a small rural bias in terms of population shares. Yet, in 2004, about 90 percent of the poor live in the rural area so in that sense one may say there is an urban bias in the benefits. The split also shows that the average beneficiary in the urban area gets twice the support of an average rural beneficiary. This reflects that social insurance benefits and education fee exemptions are higher in the urban areas. However, when looking at the joint system of direct household support from the government the alleged division of benefits into a minority group in the formal (urban) sector with generous support and a majority (rural) group with insufficient support, is no longer evident. Once more, this does not imply that the targeting of government support cannot, and should not, be improved.

#### 5. Social Transfers and Poverty 2002-2004

Up to this point the distinction between poor and non-poor individuals has only been made implicitly by noting that the poorest quintile in the distribution of actual consumption expenditure corresponds almost exactly to the individuals below the poverty line. In this section we take a more direct approach by looking at the impact of social transfers on changes in poverty from 2002 to 2004 using the official GSO poverty line to identify the poor.<sup>94</sup> The way we measure the impact of social transfers on poverty is based on a suggestion in Ravallion, van de Walle and Gautam (1995) which was applied to Vietnamese data by van de Walle (2004a,b). The idea is basically to compare actual poverty levels and transition matrices to counterfactual poverty levels and transition matrices. In this section we only look at one counterfactual by simulating the poverty situation in Vietnam in 2002 and 2004 without social transfers.

In the estimation of poverty transition matrices we use the panel data part of the 2002 and the 2004 VHLSS, which is the same data we used in the regressions in Section 3. To verify that this sub-sample of the two surveys is representative for the Vietnamese population with respect to the incidence of poverty we first compare the official poverty incidence estimates with the estimates obtained when using the panel data and the 2002 VHLSS sampling weights. This is done in Table 12 and, as seen, the differences between the official poverty incidence estimates and the sub-sample estimates are, in general, reasonably small, and they are statistically insignificant when considering the sampling variation in the panel sample. Thus, we believe the following results are representative for the whole population.

	20	02	20	04
	Official	Panel	Official	Panel
	estimates	estimates	estimates	estimates
All Vietnam	28.9	29.5 (1.1)	19.5	20.5 (0.9)
Rural Vietnam	35.6	34.9 (1.3)	25.0	24.3 (1.1)
Urban Vietnam	6.6	7.5 (1.7)	3.6	5.1 (1.5)

Table 12: Poverty Incide	ence in Vietnam
--------------------------	-----------------

*Notes:* The official poverty incidence estimates are from GSO. The panel estimates are based on the 3935 households that were interviewed in both household surveys. The estimates are weighed using the 2002 VHLSS sample weights. Standard errors of the panel estimates, taking account of stratification and clustering, are given in parentheses. *Source:* GSO and own calculations based on the 2002 and 2004 VHLSS.

<sup>&</sup>lt;sup>94</sup> The GSO poverty line is based on the cost of basic needs approach (2100 calories per person per day). In 2004 the poverty line was 2.077 million VND (January 2004 prices).

In Table 13 we report poverty transition matrices for all of Vietnam and for the two suppopulations, Rural Vietnam and Urban Vietnam. The transition matrices on the left hand side are based on the observed per capita consumption expenditure while the transition matrices on the right hand side are based on the counterfactual measure; per capita consumption expenditure net of per capital social transfers.

		Based on observed consu expenditur in 2002 and 2	mption e 004	Based on consumption expenditure less social transfers in 2002 and 2004		
	2004			2004		
2002	Poor	Non-poor	Total	Poor	Non-poor	Total
All Vietnam						
Poor	16.5	13.0	29.5	25.4	13.9	39.3
Non-poor	4.0	66.5	70.5	5.7	55.0	60.7
Total	20.5	79.5	100.0	31.1	68.9	100.00
Rural Vietnam	1					
Poor	19.6	15.3	34.9	28.7	15.6	44.3
Non-poor	4.8	60.4	65.2	6.1	49.6	55.7
Total	24.3	75.7	100.0	34.8	65.2	100.00
Urban Vietnar	n					
Poor	3.8	3.7	7.5	11.8	7.1	18.9
Non-poor	1.3	91.2	92.5	4.4	76.8	81.1
Total	5.1	94.9	100.0	16.2	83.9	100.00

Table 13: Poverty Transitions 2002 to 2004

Notes: The poverty headcount ratios are calculated using VHLSS 2002 sampling weights.

Source: Own calculations based on 2002 and 2004 VHLSS.

In the absence of social transfers poverty would have been about 10-11 percentage points higher both in 2002 and 2004. This is a tall number. In particular, for the urban area—where actual poverty is fairly low—poverty would have been three times higher in 2004 without social transfers. In that sense, it seems fair to conclude that social transfers have had a large impact on poverty in Vietnam in 2002 and 2004. This is in stark contrast to the findings for 1993 and 1998 in van de Walle (2004b).

The above observations could have been made without the use of transition matrices. The added information in the poverty transition matrices is that it is possible to analyze if the social transfers are protecting individuals from falling into poverty or promoting individuals to escape poverty.

Table 14:	Test of	Protection	and	Promotion
-----------	---------	------------	-----	-----------

	All Vietnam	Rural Vietnam	Urban Vietnam
Protection	1.7	1.3	3.1
	(3.44)	(2.59)	(8.35)
Promotion	-0.9	-0.3	-3.4
	(1.14)	(0.33)	(6.65)

Notes: Protection is the difference between the counterfactual and actual transitions non-poor-to-poor. Promotion is the difference between the actual and the counterfactual transitions poor-to-non-poor. *z*-values are reported in parentheses. The limiting distribution of the *z*-values is the standard normal.

Source: Own calculations based on Table 3.

Ravallion, van de Walle and Gautam (1995) define protection from poverty as the difference in the movement from non-poor to poor across the actual and the counterfactual welfare measures. This movement is given as the share of the population that was non-poor in the initial year (2002) and poor in the end year (2004). In the base case (with social transfers) 4.0 percent of the population move from non-poor to poor; in the counterfactual case that share is 5.7 percent. The difference in these shares is an estimate of the protection against poverty caused by the social transfers. It simply implies that 1.7 percent of the population who are non-poor both in 2002 and 2004 would instead have fallen into poverty between 2002 and 2004 in the absence of social transfers. The estimated protection and tests of significance of the estimate are given in Table 14. As seen, social transfers have a significant protection effect. The protection effect is particularly large in the urban areas, in which we find a difference between the transition rates of 3.1 percentage points. This is a strong indication of the protection provided by the social insurance system.

Promotion out of poverty is defined analogously to protection by comparing the movement from poor to non-poor in the transition matrices. For promotion we find, in all three samples, that the movement out of poverty is larger in the counterfactual simulation than the actual case, resulting in negative promotion effects. So, fewer people are escaping poverty with social transfers than without transfers. The negative effect is, however, only significant for the urban area. The negative promotion effect is explained by the fact that many non-poor people benefit from social transfers in 2002. In the absence of social transfers they would have been below the poverty line in 2002 and, because of the large overall decrease in poverty from 2002 to 2004, a large fraction of these individuals would have escaped poverty on their own, *i.e.*, without social transfer assistance.

As the social insurance and social protection programs are aimed at protection against poverty, we conclude that the programs are reasonably successful. Obviously, this does not imply that the programs could not have been better.

#### 6. Conclusion and Policy Implementations

This chapter has updated previous analyses of the incidence of social transfers at the household level using the household survey 2004 VHLSS. The main results of the analysis paint a somewhat different picture of the distribution of social transfers than the studies based on the 1992/93 and 1997/98 household surveys. Fist of all, social transfers are not regressive in 2004. Second, the transfers had a sizeable impact on the level of poverty in both 2002 and 2004 and, furthermore, provided protection against falling into poverty in that period. The most important reason for the new results is a change in the counterfactual welfare measure used to estimate the distribution of social transfers across the welfare distribution.

In the analysis, we look at two types of transfers: social insurance and social protection, and two types of education fee exemptions: tuition fee exemptions and school contribution fee exemption. The latter are not a part of the social transfer system but we argue that education fee exemptions are an important part of direct government support to the households and that, in other fiscal systems, this kind of support would often be included in social transfers.

The social insurance system is shown to provide most support to the poorest and the richest quintiles in the welfare distribution. As such, even though it is progressive (in the relative sense) it is not pro-poor. The reason is that the social insurance system in Vietnam is not intended to be pro-poor. Starting from a government retirement system, it has evolved into a standard social insurance system providing protection for its members from age and disability related stoppage or substantial reduction of earnings. Membership of the system has expanded over the years to offer either mandatory or voluntary coverage for all formal sector workers. The main problem with the system is a low

coverage; even of formal sector workers, combined with a very large funding directly from the state budget. Therefore, the government should do more to encourage private sector employees to participate in the system, possibly by changing the benefit formula, and it should also work towards ensuring fulfillment of payment obligations from the insured members and their employers (public as well as private). At the same time the government should gradually reduce the state budget financing of the system.

The social protection program has very limited resources that should probably be increased somewhat in the future. The targeting of the social protection transfers appears to be reasonably propoor, but we are unwilling to make firm conclusions. The main problem is a difficulty in identifying the beneficiaries because the social protection program covers four distinct groups of supposed beneficiaries. Support for the extremely poor, poor people struck by natural disasters and people starving between harvests should clearly be targeted towards individuals in the lowest quintile of the welfare distribution. But beneficiaries from 'war compensation' support need not be poor as the transfers to this group of people is based on historical merits; not on a income or consumption based means testing. As we have no information about the reason for the transfer in the 2004 VHLSS we cannot separate needs tested transfers from war compensation. Since the identification problem is not only a problem in the household surveys but also causes some difficulties in understanding the public expenditure on social protection we suggest that war compensation transfers are clearly identified in the state budget (this should cause no problem as the transfers are from a special social guarantee fund), and ideally completely separated from the social protection transfers.

The education fee exemptions are also, in general, progressive, in particular the contribution fee exemptions. Also for education fee exemptions there is a mix of reasons for granting the exemption of which several are not poverty related. In the 2004 household survey we are able to identify poverty related education fee exemptions and the incidence analysis of this sub-set of exemptions shows a clear poverty orientation. The main problem with the education fee exemptions is that the monetary value of the exemptions is increasing with the welfare level. This positive association can be explained by the fact that richer households have children in more expensive schools compared to poorer households. If there is a positive relationship between price an quality, as found in Behrman and Knowles (1999), then parents of children in more expensive schools should pay for their choice of higher school quality even when they are granted education fee exemption. One solution could be to change the support system from a tax reduction to a transfer. This would allow the central government to regulate the transfer amount and, thereby, get control of the actual support to the households. In addition, this change would increase the transparency of the support system by putting the expenditure on the state budget.

#### Acknowledgements

The authors wish to thank Nguyen Thi Nguyet for inputs in the early phase of the project and Vu Hoang Dat for valuable discussions.

#### References

- Behrman, J. R., and J. C, Knowles. 1999. "Household income and child schooling in Vietnam". World Bank Economic Review, 13: 211-56.
- Demery, L. 2003. "Analyzing the incidence of public spending", in François Bourguignon and Luiz A. Pereira da Silva, eds, *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. The World Bank and Oxford University Press.
- Hadi, AS. 1992. "Identifying multiple outliers in multivariate data". Journal of the Royal Statistical

Society, Series B., 54: 761-771.

- Hadi, AS. 1994. "A modification of a method for the detection of outliers in multivariate samples". *Journal of the Royal Statistical Society, Series B.*, 56: 393-396.
- Justino, P. 2005. "Beyond HEPR: a framework for an integrated national system of social security in Viet Nam". Policy Dialogue Paper 2005/1. United Nations Development Programme. Hanoi, Vietnam.
- MOLISA. 1999. Statistics on Social Safety Net in Vietnam. Hanoi, Vietnam
- Nguyen, The Dzung. 1999. "Capacity assessment for Vietnam's National Target Program for Hunger Eradication and Poverty Reduction: Program assessment". United Nations Development Programme. Hanoi, Vietnam.
- Nguyen, Viet Cuong. 2003. :Assessing the impact of Vietnam's programs for targeted transfers to the poor using the Vietnam Household Living Standard Survey 2002". Background paper for the Vietnam Development Report 2004, prepared for the World Bank.
- Prescott, N. 1997. "Poverty, social services, and safety nets in Vietnam". World Bank Discussion Paper No. 376. The World Bank.
- Preston, D. 1999. "Social Safety Nets in Vietnam". International Labour Organisation. Processed
- UNDP-MOLISA. 2004. Taking Stock, Planning Ahead: Evaluation of the National Targeted Programme on Hunger Eradication and Poverty Reduction and Programme 135. United Nations Development Programme. Hanoi, Viet Nam.
- Ravallion, M., D. van de Walle, and M. Gautam. 1995. "Testing a social safety net". *Journal of Public Economics*, 57: 175-99.
- Shaffer, P., and Thang, Nguyen. 2004. "Study of Targeting Efficiency of NTP on HEPR". Consultant report for UNDP and MOLISA. Hanoi, Vietnam.
- SRV. 2003. National Education for All (EFA) Action Plan 2003 2015. Hanoi, Vietnam.
- SRV. 2005. Vietnam. Managing Public Expenditure for Poverty Reduction and growth. Public Expenditure Review and Integrated Fiduciary Assessment. A joint Report of the Socialist republic of Vietnam and the World Bank prepared with the support of the Likeminded Donor Group. Financial Publishing House. Hanoi.
- van de Walle, D. 2003. "Behavioral Incidence Analysis of Public Spending and Social Programs", in François Bourguignon and Luiz A. Pereira da Silva (Eds.) *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools.* The World Bank and Oxford University Press.
- van de Walle, D. 2004a. "Testing Vietnam's public safety net". *Journal of Comparative Economics*, 32: 661-79.
- van de Walle, D. 2004b. "The static and dynamic incidence of Vietnam's public safety net". in Paul Glewwe, Nisha Agrawal and David Dollar, eds, *Economic Growth, Poverty and Household Welfare: Policy Lessons from Vietnam*. World Bank Regional and Sectoral Studies. World Bank, Washington, DC.
- World Bank. 2002. "Localizing MDGs for poverty reduction in Viet Nam: reducing vulnerability and providing social protection". Poverty Task Force. World Bank Vietnam.

World Bank. 2004. Vietnam Development report 2005: Governance. Hanoi, Vietnam.

World Bank. 2005. Vietnam Development report 2006: Business. Hanoi, Vietnam.
### **Appendix 1. Regulation on Social Allowances**

In this appendix we list the precise descriptions of the target beneficiaries and the transfer amounts as they are given from the Government of Vietnam. The current regulations on social allowances are presented in Decree 07/2000/ND-CP of the Government. The regulations on the allowances and subsidies to those contributed to Vietnam's revolutions are stated in Decree 210/2004/ND-CP of the Government.

From Decree 07/2000/ND-CP notices can be drawn as follows:

### I. Regular Social Allowances

### A - Targeted beneficiaries:

1. Orphan children aged less than 16 and children who cannot survive on their own.

2. Elderly people aged above 60 who live alone and old people who do not have relatives to rely on and who cannot live on their own.

- 3. Severe invalids who do not have income or who cannot live on their own.
- 4. Mental patients who live alone or in poor families that cannot afford the living costs.

### *B* – *Regulations on the transfer amount:*

1. The minimum amount the beneficiaries are given by commune authority is VND 45,000 per person per month.

2. In State social institutions the minimum amount is VND100,000/person/month. For children less than 18 months of age the minimum amount of VND 150,000/child/month.

### II. Irregular Social Allowances (once)

#### A - Targeted beneficiaries:

Those who fall into difficult situations due to unexpected disasters or other unforeseen events can be given social allowances. Concretely, those people include:

- 1. Households who have family member(s) that died or is missing.
- 2. Households whose houses are heavily damaged (corrupted, fired...).
- 3. Households who lose production equipment and fall into hunger.
- 4. People who are heavily injured.
- 5. People who live in hunger in the between-crop periods.
- 6. Homeless beggars who are planned to return home.
- B Transfer amount:

The amount of irregular social allowance is decided by the provincial Peoples Committee and may vary.

# Regulations on Allowances and Subsidies to Those Contributed to Vietnam Revolutions (Decree No. 210/2004/ND-CP)

Details about the benefit rates are presented in the appendices of the Decree as follows:

### Annex 1: Monthly Rates of Allowances for Those Who Contributed to Vietnam Revolutions

No **Targeted beneficiaries** Regular Minor allowances allowances 01 People took part in the revolution before 1945 250 - Migrant 60/1 year old - Non-migrant 540 - Family relatives of those (former soldiers) who took part in the revolution before 1945 receive monthly allowances when the former soldiers die. 292 - Those who are family relatives of those (former soldiers) who took part in the revolution before 1945 and live alone receive monthly allowances when the 495 former soldiers die. 02 - Those who participated in the early periods of the Revolution August 1945. 292 03 Relatives of martvr - Death allowances for relatives of one martyr 292 495 - Death allowances for relatives of more than one martyr - Death allowances for relatives of a martyr who have no one to lean on 495 04 Vietnam Mother heroes 710 05 Military heroes, labor heroes during the war periods 250 06 - Invalid former soldiers and people who are treated as invalid former soldiers + Those who lose more than 81% working capacity 150 + Those that lose more than 81% working capacity and bear a heavy wound 292 - Those who serve invalid soldiers, people treated as invalid + Those who lose more than 81% working capacity 292 + Those that lose more than 81% working capacity and bear a heavy wound 380 - Relative of invalid soldiers, people treated as invalid soldiers receive monthly allowance when the soldiers die 175 - Relatives of invalid soldiers, people treated as invalid soldiers who live alone receive monthly allowances when the soldiers die 390 07 - Soldiers bear a job accident + Those who lose more than 81% working capacity 150 + Those who lose more than 81% working capacity and bear a heavy wound. 292 - Those who servers soldiers sick due to a job accident + Those who lose more than 81% working capacity 292 + Those who lose more than 81% working capacity and bear a heavy wound. 380 - Relatives of soldiers sick (lose more than 61% working capacity) due to a job accident receive monthly allowances when the soldiers die 175 - Relatives of soldiers sick (lose more than 61% working capacity) due to a job accident receive monthly when the soldiers die and they live alone 390 08 - Soldier patients + Those who lose more than 81% working capacity 150 + Those that lose more than 81% working capacity and bear a heavy wound. 292 + Those who lose 61% - 70% working capacity 489 + Those who lose 71% - 80% working capacity 564

677

+ Those who lose 81% - 90% working capacity

Measurement unit: VND thousand

## MARKET, POLICY AND POVERTY REDUCTION IN VIETNAM

No	Targeted beneficiaries	Regular allowances	Minor allowances
	+ Those who lose 91% - 100% working capacity	752	
	- Those who serve soldier patients		
	+ Those who lose more than 81% working capacity	292	
	+ Those who lose more than 81% working capacity and bear a heavy wound.	380	
	- Relatives of sick soldiers receive monthly allowances when the soldiers die	175	
	- Relatives of sick soldiers who live alone receive monthly allowances when the	9	
	soldiers die	390	
09	Soldiers who are sick due to the job		
	+ Those who lose 41% - 50% working capacity	300	
	+ Those who lose 51% - 60% working capacity	376	
10	- Those who assisted the Revolution before August 1945		
	+ Monthly allowances	292	
	+ Allowances for lonely people	495	
	- Those who assisted the revolutions during the war		
	+ Monthly allowances	175	
	+ Allowances for lonely people, solitaries	390	
11	Monthly educational allowances	292	

Annex 2 Monthly Allowances for Invalid Soldiers and People Treated as Invalid Soldiers

No	Working capacity loss - %	Rates of allowances	No	Working capacity loss - %	Rates of allowances
01	21%	197	41	61%	573
02	22%	207	42	62%	583
03	23%	216	43	63%	592
04	24%	226	44	64%	602
05	25%	235	45	65%	611
06	26%	244	46	66%	620
07	27%	254	47	67%	630
08	28%	263	48	68%	639
09	29%	273	49	69%	649
10	30%	282	50	70%	658
11	31%	291	51	71%	667
12	32%	301	52	72%	677
13	33%	310	53	73%	686
14	34%	320	54	74%	696
15	35%	329	55	75%	705
16	36%	338	46	76%	714
17	37%	348	57	77%	724
18	38%	357	58	78%	733
19	39%	367	59	79%	743
20	40%	376	60	80%	752
21	41%	385	61	81%	761
22	42%	395	62	82%	771
23	43%	404	63	83%	780
24	44%	414	64	84%	790
25	45%	423	65	85%	799
26	46%	432	66	86%	808
27	47%	442	67	87%	818
28	48%	451	68	88%	827
29	49%	461	69	89%	837
30	50%	470	70	90%	846
31	51%	479	71	91%	855
32	52%	489	72	92%	865
33	53%	498	73	93%	874
34	54%	508	74	94%	884
35	55%	517	75	95%	893
36	56%	526	76	96%	902
37	57%	536	77	97%	912
38	58%	545	78	98%	921
39	59%	555	79	99%	930
40	60%	564	80	100%	940

Measurement unit: VND thousand

Annex 3
Monthly Allowances for Soldiers Who Are Sick due to a Job Accident

Measurement unit: VND thousand

No	Working capacity loss - %	Rates of	No	Working capacity loss - %	Rates of
		allowances		<b>5</b>   <i>y</i>	allowances
01	21%	158	41	61%	459
02	22%	165	42	62%	466
03	23%	173	43	63%	474
04	24%	180	44	64%	481
05	25%	188	45	65%	489
06	26%	195	46	66%	496
07	27%	203	47	67%	504
08	28%	210	48	68%	511
09	29%	218	49	69%	519
10	30%	225	50	70%	526
11	31%	233	51	71%	534
12	32%	240	52	72%	541
13	33%	248	53	73%	549
14	34%	255	54	74%	556
15	35%	263	55	75%	564
16	36%	271	46	76%	571
17	37%	278	57	77%	579
18	38%	286	58	78%	586
19	39%	293	59	79%	594
20	40%	301	60	80%	602
21	41%	308	61	81%	609
22	42%	316	62	82%	617
23	43%	323	63	83%	624
24	44%	331	64	84%	632
25	45%	338	65	85%	639
26	46%	346	66	86%	647
27	47%	353	67	87%	654
28	48%	361	68	88%	662
29	49%	368	69	89%	669
30	50%	376	70	90%	677
31	51%	383	71	91%	684
32	52%	391	72	92%	692
33	53%	398	73	93%	699
34	54%	406	74	94%	707
35	55%	414	75	95%	714
36	56%	421	76	96%	722
37	57%	429	77	97%	729
38	58%	436	78	98%	737
39	59%	444	79	99%	744
40	60%	451	80	100%	752

## **Appendix 2. Summary Statistics**

	Overall	Overall	Mean of	Sd. of	Number of	Minimum	Maximum
	Mean	Standard	non-zero	non-zero	non-zero		
Variable		deviation	observations	observations	observations		
Real household expenditure	3,320	10,177	3,320	10,177	3,935	-154,481	108,954
Social insurance benefits	252	2,349	1,804	6,057	539	-50,529	50,443
Social protection benefits	18	700	141	1,950	516	-13,996	7,817
Sum social insurance							
and social protection	270	2,344	1,172	4,771	912	-50,529	50,443
Tuition fee exemption	228	4,430	3,952	18.078	217	-70,307	107,537
Contribution fee exemption	126	2,644	1,454	8.892	362	-62,060	65,946
Sum of fee exemptions	354	5,691	3,052	16.483	470	-132,367	107,537
Household size	-0.103	1.202	-0.254	1.883	1590	-10	7
Gender of household head	0.013	0.233	0.241	0.973	214	-1	1
Children 0-5	-0.032	0.541	-0.142	1.123	921	-3	3
Children 6-10	-0.086	0.593	-0.283	1.049	1214	-3	3
Primary-educated	-0.020	1.118	-0.038	1.534	2106	-7	6
Secondary-educated	-0.018	0.978	-0.042	1.500	1668	-5	5
High-school-educated	-0.004	0.654	-0.017	1.324	924	-4	5
Vocationally educated	0.082	0.418	0.788	1.059	410	-2	4
Professionally educated	0.031	0.382	0.318	1.185	379	-4	2
University-educated	0.020	0.323	0.290	1.191	260	-3	3
Females above age 55	0.022	0.322	0.231	1.025	366	-2	2
Males above age 60	0.010	0.249	0.164	0.989	243	-1	1

### Table A: Summary Statistics for the Variables in the Regressions in Section 3.

Notes: All variables are measured as changes from 2002 to 2004. Monetary values are measured in 1,000 VND, January 2004 prices. The means and standard deviations are weighted using the sample inflation factors from 2002 VHLSS.

# 5

# Ethnicity and Poverty Reduction

Hoang Thi Thanh Huong, Pham Hương Giang, Tran Binh Minh and Henrik Hansen

### 1. Introduction

Vietnam is a multi-ethnic country with 54 ethnic groups coexisting. The Kinh people are by far the largest group accounting for 86 percent of the population. Among the ethnic minorities the most populous are Tay, Thai, Muong, and Khmer with populations of around one million each, while the least populous are Brau and Odu each with less than a thousand people. The Hoa (Chinese) was the second largest ethnic group in the population census of 1979 but moved to the 6<sup>th</sup> place 20 years later in the 1999 Population and Housing Census (GSO 2001).

The Kinh people mainly inhabit the low land and the two deltas (the Red River Delta and the Mekong River Delta). Most of the other ethnic groups are scattered over mountainous areas and the midland spreading from the North to the South. Some ethnic groups are highly concentrated in the plains, though, such as the Khmer in the Mekong River Delta, the Hoa in Ho Chi Minh City and the Mekong River Delta and the Cham on the Southern coasts. All ethnic groups have their own diverse and unique cultures.

At the national level Vietnam has made impressive achievements in terms of economic growth and poverty reduction. The national poverty rate has been reduced greatly, from 58.1 percent in 1993 to 19.5 percent in 2004. However, the benefits of the remarkable success are not evenly distributed across ethnic groups. The Kinh and Hoa households have made largest progress with a poverty rate declining from 53.9 percent in 1993 to 13.5 percent in 2004, but the other ethnic groups (the minorities) are lagging behind in poverty reduction. In 1993 the poverty rate among the minorities was 86.4 percent, dropping to 60.7 percent in 2004 (World Bank 2003 and CAF 2006).

In this chapter we look at some explanations for the widening gap in the poverty rates between the majority group (the Kinh and the Hoa) and the ethnic minorities using data from three nationally representative household surveys. In previous studies of the gap in living standards between the Kinh-Hoa majority and the ethnic minorities it is generally concluded that the higher incidence of poverty among the minorities is caused both by their lower endowments, which are systematically different from the Kinh-Hoa's, and by systematically lower gains from the productive endowments and adverse geographical factors.

The first study that seeks to decompose the gap in living standards into differences in endowments and differences in returns to the endowments by van de Walle and Gunewardena (2001) is based on

data from the 1993 Vietnam Living Standard Survey (1993 VLSS). Based on regression analyses and a Blinder-Oaxaca decomposition of the average per capita consumption expenditures at the household level, this study finds that fundamentally different models generate incomes for the majority and the minorities. Moreover, the differences in returns to productive characteristics explain a very large fraction of the ethnic inequality and this is most pronounced in areas where both groups live.<sup>95</sup> In addition to the differences in returns the authors also find strong geographical effects and they conclude that there is an important role for geographically targeted programs, which should be tailored towards the minorities in order to take account of the differences in the income generating models between the majority and the minority groups.

A second study by Baulch et al. (2004) examines the quantitative evidence on disparities in living standards between and among different ethnic groups in Vietnam. Using data from the 1998 Vietnam Living Standard Survey (1998 VLSS) and the 1999 Population and Housing Census, Baulch et al. show that the Kinh and the Hoa have substantially higher living standards compared to other ethnic groups. Subdividing the population into five broad categories the authors find that while the Kinh, the Hoa, the Khmer, and the Northern Highland minorities have benefited from economic growth in the 1990s the growth of the Central Highland minorities has stagnated. Disaggregating further, they also find that the ethnic groups whose living standards have risen fastest are those that have the highest school enrollment rates, are most likely to intermarry with Kinh partners, and are the least likely to practice a religion. While some ethnic minorities seem to be doing well from a strategy of assimilating (both culturally and economically) with the Kinh-Hoa majority, other groups are attempting to integrate economically while retaining distinct cultural identities. A third group comprising the Central Highland minorities, including the Hmong, is largely being left behind by the growth process. In addition to the detailed descriptive analyses, Baulch et al. also update, and modify the decomposition analysis by van de Walle and Gunewardena and show that even if minority households had the same endowments as Kinh households, this would close no more than a third of the gap in average per capita consumption expenditures. Hence, there still appear to be important structural differences in the income generation models in 1998.

The rapid growth in Vietnam, the associated structural changes in the economy and the many recent programs and policies initiated to help ethnic minorities may have changed the situation for the minority groups radically since 1998. However, the only recent study, by Nguyen (2006), is fairly brief. Therefore, there is a need for an updated, in depth, study of the ethnic minority poverty relative to the majority in order to assist the Government of Vietnam in designing effective programs for poverty reduction among the ethnic minorities. Accordingly, this chapter provides an update of the analyses by van de Walle and Gunewardena (2001) and Baulch *et al.* (2004) in which household level data from the three surveys 1993 VLSS, 1998 VLSS and 2004 VHLSS are used and compared.

The chapter is organized as follows. In Section 2 some government policies and programs aiming at ethnic minorities are briefly reviewed. Next, in Section 3, a range of socio-economic indicators are used to examine the differences in living standards of the minorities in a comparison with the majority. In Section 4 we present and discuss regressions and decompositions of the gap between the consumption expenditures in majority and minority households. Finally, Section 5 provides a brief summary and offers some concluding remarks, policy recommendations and suggestions for future research.

### 2. Government Policies towards Ethnic Minorities

The Government of Vietnam has realized the importance of improving the living standards of the

<sup>&</sup>lt;sup>95</sup> Nguyen (2006) has updated the analysis by van de Walle and Gunewardena to cover all four nationally representative household surveys and she finds fairly small changes over time.

ethnic minorities and reducing the income gaps between different ethnic groups and therefore developed a number of supportive policies and programs. In general, the ethnic minorities can benefit from (i) general socio-economic development programs such as the Comprehensive Poverty Reduction and Growth Strategy, (ii) national targeted programs such as Program 135 and the national program on Hunger Eradication and Poverty Reduction, and (iii) policies and programs directly targeted at the ethnic minorities.<sup>96</sup>

The most important of the general programs is the Comprehensive Poverty Reduction and Growth Strategy (CPRGS), which was approved by the Prime Minister in 2002. The CPRGS is a general action plan to achieve economic growth and poverty reduction objectives translating the Government's Ten-Year Socio-Economic Development Strategy, the Five-Year Socio-Economic Development Plan as well as other sectoral development plans into concrete measures with well-defined road maps for implementation. The tasks and objectives contained in the CPRGS not only calls for targeted measures to support specific poor groups but also sees linkages within the matrix of policies that must work in tandem to ensure sustainable development. It includes macroeconomic policies, policies on structural adjustment, sectoral development policies and measures, and social welfare policies of all sectors and at all levels. Among the specific objectives set out in the CPRGS some are targeted directly to extremely disadvantaged groups and ethnicities including provision of basic infrastructure, improvement of living standards of ethnic minorities and preservation and development of ethnic minority cultures.

To concretize objectives towards ethnicity, the social and economic development program for communes in extreme difficulties in ethnic and mountainous areas, which is known as Program 135, was launched in 1998. Program 135 focuses on the mountainous and remote areas with the main objective of developing infrastructure in the villages and communes facing extreme difficulties as well as developing inter-communal infrastructure (roads, small irrigation dams, clean water system, schools, health centers and other infrastructure). Initially the Program covered 1715 communes, of which 1568 were mountainous and 147 were in lowland areas, with a population of around 1.1 million households and over 6 million people. It has recently been expanded to cover 2362 poor and remote communes.

With the target group of poor households, the main objectives of the National Program on Hunger Eradication and Poverty Reduction (HEPR) was initially to eliminate chronic hunger, to reduce the national poverty rate and to narrow down the gap between the disadvantaged and other regions in the country through enhancing the socio-economic development (SRV 2001). In 2001-2005 the HEPR program focused on providing credit for poor households to develop production, guiding poor people on business, agricultural and forestry extension, developing models of hunger eradication and poverty reduction in poor communes, supporting infrastructure in about 700 poor communes (those not under Program 135), assisting production in terms of processing of agricultural production, non-agricultural skill development, small-scale industries and services and capacity building through training for cadres and official at the communes level.

Program 135 and the HEPR program were evaluated in 2004 (see UNDP and MOLISA 2004). The evaluation concludes that each of the sub-programs in the HEPR, namely credit, health and tuition exemptions are reasonably well targeted. However, while 70 percent of the benefits of each of these sub-programs reach the poor their coverage of the total poor households remains limited. Program 135, using rather simple and straightforward funding approaches, covers nearly a quarter of all communes, thereby reaching about 15 percent of the total population. As more than half of these households are below the poverty line, Program 135 is fairly well targeted. Moreover, support for health care, tuition waivers and infrastructure have a considerably greater positive impact on the poor than credit and farming extension services (UNDP and MOLISA 2004).

<sup>&</sup>lt;sup>96</sup> A chronological list of directly targeted policies and initiatives are given in the Appendix.

In addition to the national target programs, production supportive programs for ethnic areas have been implemented. The Government has launched many policies supporting the production by those people such as resettlement to develop new economic areas, policies on forest allocation, land allocation and land reclamation polices. In particular, production land and supportive policies are targeted to ethnic minority people in the Central Highlands and land re-acquisition for Khmer people in the Mekong River Delta. Under these policies, land is allocated to landless people and forest is handed over partially to stabilize the living standards of ethnic minorities and partially to protect and restore natural resources in those areas.

The Government also pays attention to education and training in the ethnic and mountainous areas. Many preferential policies have been promulgated, including those encouraging teachers to work in extremely difficult and remote areas, scholarships for ethnic pupils and students, granting paper and textbooks for ethnic pupils and teaching ethnic languages to officials and public servants working in ethnic minority and mountainous areas. Policies on rotating civil servants and training for staff have also been implemented, increasing the number of ethnic minority officers at all levels and improving the qualification and expertise of those people. In terms of healthcare there are many preferential policies for ethnic minority and mountainous areas including reductions and exemptions of hospital fees for ethnic minority people and diagnosis and treatment for poor patients.

In general, many of the implemented policies aim at creating a framework for poverty reduction and socio-economic development making significant progress towards the goals of equality and parity among ethnic minority and other regions in the country. However, there is a concern that a large number of policies for ethnic minorities are not fully consistent with each other (Baulch *et al.* 2004). Hence, in order to assist the Government of Vietnam in formulating effective policies for socio-economic development among the ethnic minorities it is necessary to analyze the impact of past policies much as it is done in UNDP and MOLISA (2004). Unfortunately, the many different policy initiatives makes it difficult to apply the more advanced statistical tools developed for impact evaluation.<sup>97</sup> Therefore, we follow the approach by van de Walle and Gunewardena (2001) and Baulch *et al.* (2004) and look more broadly at the outcome of the policies in terms of the socioeconomic development and the differences in living conditions across the ethnic groups in the rural areas of Vietnam.

### 3. The Socio-Economic Development of the Ethnic Groups

The socio-economic development of some broadly defined ethnic groups can be investigated using the Vietnamese living standard surveys. Since 1993 Vietnam has conducted four nationally representative living standard surveys (1993 VLSS, 1998 VLSS, 2002 VHLSS and 2004 VHLSS). In this chapter we use three of these surveys to describe the differences in the socio-economic development among the ethnic groups.

The development in poverty and other socio-economic measures differs markedly between the urban and rural areas of Vietnam from 1993 and onwards. As most of the ethnic minority groups are located in the rural areas a comparison of the minority groups and the majority (Kinh-Hoa) that includes the urban areas would be misleading. Therefore, we focus on the rural areas of Vietnam using the definition of rural and urban areas provided in the household surveys. Limiting the coverage to the rural areas decreases the sample sizes but we believe the loss in observations, which is almost exclusively for the majority, is outweighed by the increased similarity of the populations we investigate.

<sup>&</sup>lt;sup>97</sup> See Bourguinon and Silva (2003) for a discussion of different statistical techniques for impact evaluation of economic policies.

In the 1993 VLSS there are 3294 rural households in which the household head belongs to the majority ethnic group (Kinh-Hoa) while there are 545 rural households with a minority household head.<sup>98</sup> The number of observations in the 1998 VLSS is slightly larger with 3589 majority and 680 minority rural households. Finally, the 2004 VHLSS has 5620 majority and 1198 minority households in the rural areas. Given the small samples (and small populations) for many of the minority groups we cannot disaggregate the minorities beyond fairly broad groups. Therefore, the main part of the analysis just divides the total rural sample of households into a majority group (Kinh and Hoa) and a minority group consisting of the other 52 ethnicities.<sup>99</sup>

### 3.1. Poverty and Per Capita Consumption Expenditure

In Table 1 we show the estimated incidences of poverty for the rural population with a breakdown into the majority and minority groups. In addition, we further divide the minorities into some of the larger minority groupings; Northern mountainous area minorities, Central Highlands minorities and the Khmer (who are mainly in the Mekong River Delta).<sup>100</sup> The estimated poverty incidences are based on the GSO poverty line whereby an individual is poor if he or she lives in a household in which the annual per capita real consumption expenditure in January 1998 prices is below VND 1.790 million.

	1000	1000	0004	Change 1993-	Change 1998-
	1993	1998	2004	1998	2004
Rural Vietnam	63	46	24	-17	-22
Kinh-Hoa majority	60	39	16	-21	-23
Minorities	84	76	58	-8	-18
Northern Mountainous areas minorities	84	75	57	-9	-18
Central Highlands minorities	94	91	74	-3	-17
Khmer	75	61	37	-14	-24

Table 1: Incidence of Poverty in Rural Vietnam 1993, 1998 and 2004 (%)

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

The incidence of poverty in the rural area was 63 percent in 1993 dropping to 46 percent in 1998 and further to 24 percent in 2004. Throughout the period the poverty incidence for the Kinh-Hoa majority was lower than the overall average, while the poverty among the minority groups was consistently higher than the overall average. Focusing on the changes over time it is clear that the poverty reduction, measured as the change in the poverty incidence, was larger in the second half of the period from 1998 to 2004. This trend is common for all ethnic groups in Table 1. For the minorities as a whole the decline in poverty from 1993 to 1998 was very modest compared to the decline in Kinh-Hoa poverty. In particular, for the minority groups in the Northern mountainous areas and in the Central Highlands the poverty incidence fell only very little from 1993 to 1998. But in the second part, the reduction in poverty is almost as big as for the majority. Given the substantial sampling uncertainty

<sup>&</sup>lt;sup>98</sup> Using the household surveys, we have to categorize the ethnic groups according to the ethnicity of the household head even though he or she may be married to an individual from another ethnic group. The extent of inter-marriages is quite limited in Vietnam, though. See Baluch *et al.* (2004).

<sup>&</sup>lt;sup>99</sup> There has been some discussion of whether the Hoa should be analyzed separately rather than grouped with the Kinh majority. The reason is that the distribution of per capita expenditure for the Hoa is bi-modal in 1998 with the urban Hoa people being generally richer than the Kinh people (Baulch *et al.* 2004). However, as we focus on the rural areas we use the 'common' grouping into the majority (Kinh and Hoa) and the other ethnic groups as the minority.

<sup>&</sup>lt;sup>100</sup> The grouping of the ethnicities is given in Table A-1 in the Appendix.

we cannot reject the hypothesis that the drop in the poverty rates is the same for all groups from 1998 to 2004. This implies that, in terms of poverty *changes*, the minorities are now taking part in the overall progress in the Vietnamese economy. However, considering the high level of poverty among the minority groups compared to the Kinh-Hoa and the many policy initiatives discussed in Section 2, the relevant question is why the poverty rates did not decrease more for the ethnic minorities than for the majority group.

Table 1 also shows a large variation within the ethnic minority groups. The Khmer (in the south of Vietnam) has a much lower poverty incidence than the minority groups in the Central Highlands and in the north of Vietnam throughout the decade. This may possibly be explained by the generally lower poverty incidence in the south of Vietnam compared to the centre and the north. Hence, it may be that the Khmer poverty relative to the southern Kinh-Hoa poverty is comparable to the relative poverty between the majority and the minorities in the Central Highlands and in the Northern mountainous area. This illustrates the main problem in the analyses of ethnic well being and development in Vietnam, namely that it is very difficult to separate ethnicity and location.

In Figure 1 and Table 2 we move beyond the poverty incidence measures and look at per capita real consumption expenditure. The left column of graphs in Figure 1 plots the cumulative distribution functions of per capita real consumption expenditure for the Kinh-Hoa and the minority groups. In all three years there is first order stochastic dominance as the cumulative distribution function for the Kinh-Hoa is everywhere below the cumulative distribution function for the minority groups. Hence, the higher incidence of poverty among the minority groups is not dependent on the specific choice of poverty line. All measures of poverty that are decreasing in per capita consumption expenditure will be lower for the Kinh-Hoa majority group than for the ethnic minority group.

In Table 2 we report the average per capita consumption expenditure estimates for the ethnic groups. Already in 1993 the average per capita real expenditure for the Kinh-Hoa majority was above the GSO poverty line.<sup>101</sup> In contrast, the average per capita real expenditure was well below the poverty line for the three minority groupings. Moreover, for the Central Highlands minorities the average expenditure was still below the poverty line in 2004.

<sup>&</sup>lt;sup>101</sup> But the median per capita real expenditure was below the poverty line as the poverty incidence was 63 percent.



Figure 1: Cumulative Distribution of Real Per Capita Expenditure 1993, 1998 and 2004

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

	1993	1998	2004	Average annual growth rate 1993-1998	Average annual growth rate 1998-2004
Kinh-Hoa	1822.9	2311.5	3188.8	4.9	5.5
Minorities	1260.2	1503.7	1869.4	3.6	3.7
Northern mountainous minorities	1313.8	1560.1	1878.8	3.5	3.1
Central Highlands minorities	951.4	1080.7	1412.8	2.6	4.6
Khmer	1409.9	1839.7	2516.8	5.5	5.4

# Table 2: Per Capita Real Consumption Expenditure in Rural Vietnam 1993,1998 and 2004

*Note:* Per capita real consumption expenditure is measured in thousand VND, January 1998 prices. The annual average growth rates are in percent.

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

Looking at the growth rate in the average per capita expenditure, we find the Khmer having the highest overall growth in expenditure followed by the Kinh-Hoa. For both groups the growth in real expenditure is substantial, exceeding five percent per year. The two other minority groups have lower growth rates and for the Northern mountainous minorities the growth appears to be lower in the second period compared to the first.<sup>102</sup> The main result in Table 2 is that the average expenditure levels across the ethnic groups are diverging as the growth rate is higher for the Kinh-Hoa majority compared to the minorities, taken together. In the following we seek to provide some explanations of the reasons why the ethnic minority groups are poorer than the majority and why the expenditure gap seems to be widening despite the Government of Vietnam's efforts to increase the welfare of the ethnic minorities.

#### 3.2. Household Characteristics and Endowments

Even when the population of households is limited to the rural areas of Vietnam we record significant differences in household composition and endowments between the Kinh-Hoa majority and the minorities. Table 3 reports some of these differences in the form of estimated population means for the two groups in the years 1993, 1998 and 2004.

The average minority household is larger than the average majority household in terms of the number of individuals in the households (household size) in all three years. For both groups the average household size has decreased from 1998 to 2004. In 1998, both the typical (e.g., the median) Kinh-Hoa household and the typical minority household has five members. In 2004, the typical Kinh-Hoa household has only four members while the typical minority household still has five members. Hence, for the minority households the fall in the average household size reflects a drop in the frequency of large (6+) households.

<sup>&</sup>lt;sup>102</sup> Here it is important to be aware of the sampling uncertainty in the estimates. We cannot reject the hypothesis that the growth rate in per capita real expenditure was equal in the two periods for the Northern mountainous minorities.

	19	993	1998		20	)04
	Kinh-Hoa	Minorities	Kinh-Hoa	Minorities	Kinh-Hoa	Minorities
Household size	4.88	5.54	5.43	6.11	4.25	5.28
Proportion of household that is						
Ages 0 to 6	0.17	0.21	0.12	0.15	0.08	0.11
Ages 7 to 16	0.23	0.23	0.27	0.29	0.21	0.26
Male over age 16	0.28	0.27	0.29	0.26	0.33	0.31
Female over age 16	0.34	0.30	0.32	0.29	0.38	0.32
Proportion of household consisting of						
One or two adults	0.07	0.03	0.03	0.01	0.10	0.03
Parent(s) and one child	0.13	0.10	0.06	0.03	0.14	0.08
Parent(s) and two children	0.18	0.13	0.17	0.12	0.28	0.22
Parent(s) and three or more children	0.37	0.40	0.43	0.42	0.25	0.35
Three-generation household	0.03	0.04	0.07	0.09	0.16	0.19
Other	0.22	0.30	0.25	0.33	0.08	0.13
Age of head of household, years	45	42	47	44	50	45
Education level of head of household						
< primary school	0.37	0.51	0.36	0.51	0.29	0.47
Primary school	0.25	0.23	0.24	0.23	0.26	0.26
Lower secondary	0.25	0.17	0.25	0.15	0.28	0.19
Upper secondary school	0.05	0.02	0.05	0.02	0.07	0.02
Short-term technical worker	0.04	0.04	0.04	0.04	0.03	0.02
Long-term technical worker & prof.sec. school	0.04	0.03	0.05	0.03	0.06	0.04
Bachelors and above	0.01	0.01	0.01	0.01	0.01	<0.00
Proportion of members	0.46	0.28	0.52	0.35	0.67	0.47
that completed primary school and above						
Land area per household (square meters)						
Annual land	3866	5663	4734	4770	3530	7095
Perennial land	669	956	1447	1637	1114	1728
Forestry land	110	59	524	4553	588	5062
Water surface land	170	1109	219	71	399	112
Other land	118	498	201	842	570	868
Proportion of households receiving remittances	0.21	0.12	0.21	0.09	0.89	0.79
Proportion of members working	0.13	0.02	0.16	0.03	0.20	0.05
for Non farm self employment						

Table 3: Composition and Endowments of Kinh-Hoa and Minority Households in Rural Vietnam, 1993, 1998 and 2004

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

Next, the composition of the households in terms of children and adults shows that majority households have fewer small children, on average, compared to minority households. Moreover, the proportion of small children is clearly decreasing over time for both groups. Measured by the change in percentage points the decrease in the number of small children is equal for the majority and the minorities.

The differences in household size and composition may explain a part of the expenditure gap as equivalence scales are not used in the Vietnamese household surveys. However, it is not clear that it can also explain the widening gap over time. Furthermore, a few experiments with different adult equivalence scales (not shown) reveal that the cumulative distribution function of adult equivalent consumption expenditure for the Kinh-Hoa majority dominates the cumulative distribution function for the minorities. Hence, the differences in poverty measures is not simply a result of the choice of expenditure per capita rather than per adult equivalent as the preferred welfare measure.

Turning to the endowments we start by looking at differences in human capital measured by completed formal education. A remarkable result in Table 3 is the differences in the proportions of household heads that have less than primary school (*i.e.*, no formal education). For the Kinh-Hoa majority there has been a sizable decrease in this share from 37 percent of the households in 1993 to 29 percent in 2004. In contrast, the share of minority households in which the household head has no formal education is more than half (51 percent) in 1993 and 1998 and it is still almost half of the households (47 percent) in 2004. The share of households in which the household head has completed primary school is fairly constant over the years and it is more or less the same for the majority and the minority groups (around 25 percent). Given these results it is not surprising that we find large differences in the shares of households in which the household head has completed at least lower secondary school. In terms of percentage points the largest difference is found for lower and upper secondary school. Educational levels above the upper secondary school are not wide spread in rural Vietnam neither for the Kinh-Hoa majority nor for the minorities. For both groups these households make up seven to ten percent of the population in all three years. In sum, it appears that the Kinh-Hoa majority has better human capital endowments than the minority groups and the difference has been increasing over time. Thus, education may be an important explanatory factor behind the divergence in per capita expenditure and poverty.

Another important endowment is agricultural land. In this chapter we follow the definitions in Brandt et al. (2005) and look at agricultural land for which households have long-term use rights. The land can be divided into five types; annual land, perennial land, forestry land, water surface land and other land used for agricultural production. We exclude residential land from the analysis. Table 3 reports the average land areas per household for the two ethnic groups. However, comparing the land endowments across the two groups is difficult as the distributions of the land areas are highly skewed with many small plots and a few very large plots. This partly reflects differences in the quality of land across Vietnam. Furthermore, most households only have one or two types of land whereby the averages are highly influenced by the many households that do not have the specific type of land. This is particularly important for forestry land, water surface land and other land as the median land area for these types is zero for both the majority and the minority group. Nevertheless, in terms of average land areas per household the minority have larger agricultural land areas than the majority group in all three years.

The last two rows in Table 3 are not directly parts of household composition or endowments. It is, however, information about household opportunities with respect to income generation. The second to last row reports the proportion of households receiving remittances, both domestic and foreign. There has been a significant increase in the share of households receiving remittances in the second part of the period for both groups. In 1993 and 1998, only one-in-five majority households received remittances. This share increased to almost 90 percent in 2004. Correspondingly, for the minority households about ten percent received remittances in 1993 and 1998, while this share increased to almost 80 percent in 2004. We do not record the size of the remittances but it is clear that the possibility of obtaining income from outside the commune/province to smooth consumption has increased considerably for most households.

Finally, we find that the Kinh-Hoa majority is notably more engaged in off-farm income generating activities as the share of members working for non-farm self-employment has increased from 13 percent in 1993 to 20 percent in 2004. For the minority this share was only 2 percent in 1993 and it remains low (5 percent) in 2004. Further, according to Vijverberg *et al.* (2006) only some 8 percent of the rural non-farm household enterprises are owned by ethnic minorities and Hoang (2006) finds that the productivity performance of ethnic minorities' non-farm household enterprises is much lower than that of the majority. Hence, the low entrepreneurship as well as capabilities of the ethnic minorities may be an important factor for the widening expenditure gap.

Overall, there are significant differences in endowments between the Kinh-Hoa majority and the minority groups. These differences may explain the gap in per capita expenditure. One remaining problem is to sort out if the differences in endowments are mainly related to location differences or directly to differences in ethnicity. The distinction between location and ethnicity is difficult to make as the minorities often live in remote and mountainous areas while the main part of the majority group lives in the deltas and the plains with good quality land and relatively easy access to markets.

Following van de Walle and Gunewardena (2001) and Baulch *et al.* (2004) we seek to disentangle location and ethnicity by analyzing a sub-sample of the rural population consisting of communes in which both the Kinh-Hoa majority and some of the minority groups are living. In the following, such communes are denoted 'mixed communes'. The cumulated expenditure distribution functions of per capita real expenditure for the mixed commune households are shown in the right column of Figure 1. As seen there also first order stochastic dominance in the mixed communes, so the minorities will have higher poverty regardless of the poverty measure and choice of poverty line as long as the poverty measure is decreasing in per capita consumption expenditure.

Table 4: Composition and Endowments of	f Kinh-Hoa and Minority Households in Mixed
Communes, 1	93, 1998 and 2004

	19	1993		1998		004
	Kinh-Hoa	Minorities	Kinh-Hoa	Minorities	Kinh-Hoa	Minorities
Household size	5.13	5.66	5.58	6.09	4.4	4.96
Proportion of household that is						
Ages 0 to 6	0.18	0.21	0.12	0.14	0.07	0.10
Ages 7 to 16	0.22	0.22	0.29	0.30	0.24	0.25
Male, over age 16	0.27	0.27	0.28	0.27	0.33	0.31
Female, over age 16	0.34	0.30	0.31	0.29	0.36	0.34
Proportion of household consisting of						
One or two adults	0.06	0.02	0.03	0.01	0.06	0.05
Parent(s) and one child	0.10	0.11	0.04	0.03	0.13	0.12
Parent(s) and two children	0.18	0.12	0.16	0.13	0.31	0.21
Parent(s) and three or more children	0.37	0.40	0.45	0.42	0.33	0.36
Three-generation household	0.02	0.04	0.04	0.09	0.11	0.15
Other	0.27	0.31	0.28	0.33	0.07	0.10
Age of head of household, years	44	42	45	44	47	47
Education level of head of household						
< primary school	0.40	0.44	0.37	0.46	0.24	0.40
Primary school	0.27	0.26	0.23	0.26	0.27	0.25
Lower secondary	0.20	0.19	0.26	0.18	0.29	0.24
Upper secondary school	0.04	0.02	0.04	0.02	0.07	0.03
Short-term technical worker	0.04	0.04	0.03	0.05	0.02	0.02
Long-term technical worker & prof.sec. school	0.04	0.04	0.05	0.04	0.10	0.05
Bachelors and above	<0.00	0.01	0.01	<0.00	0	<0.00
Proportion of members	0.42	0.33	0.49	0.40	0.69	0.53
that completed primary school and above						
Land area per household (square meters)						
Annual land	4996	5689	6066	5165	4274	5515
Perennial land	741	732	1381	1416	2049	2237
Forestry land	217	80	845	4961	1650	2586
Water surface land	272	1533	471	86	45	58
Other land	158	514	586	822	1083	1047
Proportion of households receiving remittances	0.17	0.14	0.17	0.10	0.86	0.85
Proportion of members working	0.13	0.03	0.14	0.03	0.19	0.11
ior non iann seir employment						

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

Estimated population means of household composition and endowments for the sub-sample of households living in mixed communes are given in Table 4. For the Kinh-Hoa majority there are only minor differences between the means of the sub-sample and the full rural sample. The most important difference is that households in the mixed communes appear to be slightly larger (more members) compared to all the rural area households. In 2004 the size difference can be traced to a higher frequency of households consisting of parents and three or more children. Another notable difference is a chance in the educational level. A larger fraction of the Kinh-Hoa household heads in the mixed communes have less than primary education in 1993 and 1998 compared to the full rural area. In 2004 we record a reversal in this result such that relatively more household heads have formal education in the mixed communes. The difference is, however, not statistically significant.

For the minorities in the mixed communes we observe smaller average household sizes compared to the full rural area and there is a smaller fraction of household heads with no formal education in all three survey years. In general, based on averages, the Kinh-Hoa majority and the minority groups in the mixed communes have more similar household compositions and endowments compared to the full rural population. This result indicates that location is a significant factor in explaining the expenditure gaps, and thus the poverty gaps, between the majority and the minorities.

It is difficult to fully demonstrate the impact of location on the socio-economic development within the majority and the minority groups. However, focusing on per capita real expenditure we can illustrate the differences between households living in mixed and non-mixed communes using the distribution of real per capita expenditure. The left column in Figure 2 shows the cumulative distribution functions for the Kinh-Hoa households living in non-mixed (pure Kinh-Hoa) communes and in mixed communes. There is no visible difference between the cumulative distribution functions for the two groups. Hence, in terms of per capita expenditure, and thus poverty, the Kinh-Hoa living in the mixed communes are not different from the Kinh-Hoa living in the non-mixed communes.

The picture is different for the minority groups. Although the cumulated distribution functions cross in 1998 and 2004, such that we do not find first order stochastic dominance, it is evident that relatively more households in the mixed communes have higher expenditure levels compared to the non-mixed (pure minority) communes. In fact, both the median and the average per capita real expenditure levels are higher in the mixed communes. In that sense the mixed commune minority groups are better off than the minority groups in the pure-minority communes. Although in no way conclusive, this indicates that location is an important factor for variation in per capita expenditure levels.

As mentioned in the introduction, the previous analyses of the majority/minority expenditure gap by van de Walle and Gunewardena (2001) and Baulch *et al.* (2004) show that differences in household characteristics and endowments alone do not explain the gaps in 1993 and 1998. Using regression analysis and a Blinder-Oaxaca decomposition, the two studies find that only about half of the gap can be explained by differences in endowments. In the next section we update the econometric analysis and the decomposition of the expenditure gap to include 2004. In doing this we use the same model specification for all three years, 1993, 1998 and 2004, to investigate if there have been significant changes in the composition of the expenditure gap over time.





Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

### 4. Econometric Analysis of the Majority/Minority Expenditure Gap

### 4.1. The Econometric Model

The econometric analysis of the expenditure gap is based on a reduced form expenditure regression model in which per capita real expenditure is regressed on the household characteristics and endowments presented in Section 3. Following standard practice, the regression model is log-linear

$$\ln(E_{hcg}) = x_{hcg}\beta_g + \eta_{cg} + \varepsilon_{hcg}$$
(1)

where  $E_{hcg}$  is the per capita real expenditure for household *h* living in commune *c* and belonging to ethnic group *g* (Kinh-Hoa or minority). The 1 x *k* vector of regressors,  $x_{hcg}$ , includes the household characteristics and endowments and  $\beta_g$  is a *k* x 1 vector of parameters (returns), which is allowed to vary across the majority and minority groups, while  $\eta_{cg}$  are commune fixed effects that are also allowed to vary across the two ethnic groups and  $\mathcal{E}_{hcg}$  are idiosyncratic random error terms.

The specification of the regression model differs slightly from the models in van de Walle and Gunewardena (2001) and Baulch *et al.* (2004). The main difference is in the specification of education. We include the highest educational level of the household head instead of the level of the most educated household member, the latter being the specification used in van de Walle and Gunewardena (2001). The main reason for this difference is that there has been an increase in the educational level of the children over time. This means that children in a household above the age of 9 or 10 will probably have completed primary school, at least in 2004, while the household head may have no formal education. We believe it should be the education of the household head that describes the level of income and consumption generating human capital which is why we use the education of the household head. Compared to Baulch *et al.* (2004) the main difference is that we include indicators of the educational level of the household head following the definition of education levels given by the Ministry of Education whereas Baulch *et al.* use a mapping to the maximum years of education of adults in the household (and its square). Hence, the Baulch et al. specification is more parsimonious than ours but the specification assumes a more restricted functional relationship between education and consumption expenditure.

Another important difference is in the specification of land endowments where both van de Walle and Gunewardena and Baulch *et al.* include measures of irrigated and non-irrigated land. We do not include such measures of land quality but only information of the type of land use as given in Tables 3 and 4.

### 4.2. Regression Results

The regression results for rural Vietnam are shown in Table 5 while the corresponding results for the sub-sample of mixed communes are given in Table 6. Table 7 reports results of Chow-tests of equality of the slope parameters across the expenditure regressions for Kinh-Hoa and the minorities for each of the three years and for the full rural sample regressions and the mixed commune regressions. Below we briefly discuss the regression results and the significance of differences in slopes.

	199	93	1998		200	)4
	Kinh-Hoa	Minority	Kinh-Hoa	Minority	Kinh-Hoa	Minority
Household size (log)	-0.324	-0.414	-0.296	-0.384	-0.217	-0.412
	(8.43)	(3.95)	(8.07)	(6.41)	(6.48)	(6.73)
Proportion of household that is						
Ages 0 to 6	-0.396	-0.211	-0.482	-0.180	-0.782	-0.647
-	(5.52)	(1.43)	(5.91)	(1.81)	(11.65)	(4.93)
Ages 7 to 16	-0.046	0.004	-0.117	0.144	-0.452	-0.319
	(0.70)	(0.03)	(2.19)	(2.30)	(8.84)	(2.78)
Males over age 16	0.210	0.392	0.155	0.186	0.163	0.059
-	(3.78)	(2.28)	(2.94)	(2.76)	(3.39)	(0.49)
Household type	. ,	. ,	ι, γ.	. ,	· · · /	. ,
Parent(s) and one child	0.157	0.223	0.104	-0.018	0.071	0.195
	(4.12)	(1.65)	(2.65)	(0.21)	(2.04)	(1.71)
Parent(s) and two children	0.200	0.165	0.099	-0.009	0.117	0.166
	(4.88)	(1.04)	(2.18)	(0.09)	(2.94)	(1.42)
Parent(s) and three or more children	0.140	0.178	0.067	-0.060	0.069	0.179
	(3.09)	(0.96)	(1.28)	(0.54)	(1.42)	(1.42)
Three-generation households	0.219	0.129	0.069	-0.048	0.092	0.215
<b>.</b>	(3.91)	(0.94)	(1.37)	(0.41)	(1.87)	(1.61)
Others	0.169	0.253	0.059	-0.003	0.047	0.127
	(3.55)	(1.33)	(1.16)	(0.02)	(0.94)	(0.95)
Age of household head	0.013	0.007	0.010	0.008	0.003	0.003
	(3.47)	(1.10)	(2.51)	(1.27)	(0.89)	(0.47)
Age of household head, squared	-0.107	-0.082	-0.080	-0.072	-0.041	-0.052
3	(3.05)	(1.14)	(2.17)	(1.09)	(1.26)	(0.91)
Education level of head of household	(0000)	()	()	()	(	(0.0.7)
Primary school	0.112	0.063	0.106	0.101	0.090	0.128
	(5.53)	(1.66)	(5.98)	(2.04)	(4.94)	(4.18)
Lower secondary	0.188	0.174	0.187	0.228	0.216	0.222
	(8.35)	(4.16)	(9.57)	(3.29)	(10.40)	(5.77)
Upper secondary	0.288	-0.009	0.310	0.197	0.272	0.200
	(7.41)	(0.07)	(11.17)	(2.56)	(9.36)	(2.82)
Short-term technical worker	0.179	0.138	0.268	0.326	0.305	0.224
	(5.15)	(2.28)	(8.00)	(4.88)	(6.55)	(3.02)
Long-term technical worker	0.338	0.285	0.319	0.396	0.460	0.441
professional secondary school	(9.47)	(2.87)	(9.72)	(5.10)	(15.80)	(5.69)
Bachelors and above	0.426	0.426	0.523	0.554	0.536	0.641
	(5.23)	(1.68)	(9.07)	(2.82)	(9.50)	(2.52)
Land area per household (ha)	()	(1100)	(0.01)	()	(0.00)	(/
Annual land	0.271	0.403	0.200	0.301	0.181	0.200
	(6.32)	(3.00)	(10.20)	(3.09)	(7.59)	(4.93)
Annual land, squared	-0.022	-0.075	-0.019	-0.037	-0.013	-0.005
· ··· ··· · · · · · · · · · · · · · ·	(4.02)	(1.52)	(7,73)	(1.31)	(3,51)	(0.92)
Perennial land	0.226	0.501	0.140	0.204	0,108	0.101
	(1.94)	(2.63)	(4.91)	(1.97)	(6.47)	(2.19)

# Table 5: Expenditure Regressions for Rural Vietnam, 1993, 1998 and 2004The Dependent Variable is Log of Real Per Capita Expenditure in January 1998 Prices

## Ethnicity and Poverty Redution

	199	1993		98	2004	
	Kinh-Hoa	Minority	Kinh-Hoa	Minority	Kinh-Hoa	Minority
Perennial land, squared	-0.048	-0.127	-0.006	-0.014	-0.005	-0.002
	(1.91)	(2.01)	(2.77)	(0.62)	(6.59)	(0.40)
Forestry land	0.454	3.941	0.141	0.057	0.034	0.047
	(2.74)	(2.77)	(2.95)	(1.52)	(1.39)	(3.53)
Forestry land, squared	-0.090	-16.65	-0.011	-0.005	-0.001	-0.002
	(2.38)	(1.60)	(1.99)	(1.01)	(1.20)	(4.19)
Water-surface land	0.117	0.176	0.124	0.956	0.110	1.290
	(1.34)	(2.54)	(1.98)	(1.50)	(1.85)	(4.34)
Water-surface land, squared	-0.032	-0.052	-0.012	-1.064	0.001	-0.396
	(1.71)	(2.26)	(1.72)	(0.76)	(0.04)	(3.25)
Other land	-0.103	0.406	0.036	0.302	0.267	0.075
	(1.63)	(1.55)	(0.74)	(2.05)	(3.28)	(0.53)
Other land, squared	0.027	-0.157	0.001	-0.145	-0.065	0.005
	(2.18)	(1.12)	(0.27)	(2.24)	(1.67)	(0.10)
Remittances	0.048	-0.027	0.075	0.070	0.046	0.086
	(2.47)	(0.62)	(4.36)	(1.55)	(1.80)	(2.27)
Non-farm self employment	0.320	0.369	0.290	0.346	0.194	0.314
	(9.85)	(3.53)	(10.53)	(3.67)	(8.76)	(3.61)
Constant	7.139	6.978	7.573	7.339	7.914	7.630
	(78.11)	(33.22)	(71.85)	(39.97)	(78.75)	(39.60)
Observations	3294	545	3589	680	5620	1198
Number of communes	115	38	130	43	1948	476
R-squared	0.26	0.36	0.29	0.38	0.28	0.39

Notes: Commune fixed effects are included in all regressions. The fixed effects sum to zero in each regression. Commune level cluster robust *t*-values in parentheses. Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

	199	93	1998		2004	
	Kinh-Hoa	Minority	Kinh-Hoa	Minority	Kinh-Hoa	Minority
Household size (log)	-0.300	-0.347	-0.248	-0.440	-0.348	-0.187
	(3.94)	(3.68)	(2.70)	(7.99)	(1.51)	(0.68)
Proportion of household of	consisting of					
Ages 0 to 6	-0.562	-0.200	-0.517	-0.258	-0.204	-1.218
	(3.21)	(0.96)	(2.72)	(1.79)	(0.30)	(1.93)
Ages 7 to 16	-0.236	-0.143	-0.115	0.170	-0.037	-0.741
	(2.20)	(0.92)	(0.89)	(1.64)	(0.08)	(2.09)
Male over age 16	0.082	0.280	0.195	0.229	0.531	0.013
	(0.71)	(1.32)	(1.55)	(2.17)	(1.14)	(0.04)
Household type						
Parent(s) and one child	0.178	0.116	0.060	0.070	0.150	0.176
	(1.43)	(0.54)	(0.66)	(0.58)	(0.60)	(0.79)
Parent(s) and two children	0.220	0.045	0.122	0.110	0.319	0.195
	(1.65)	(0.21)	(0.93)	(0.89)	(1.19)	(0.75)
Parent(s) and three or more children	0.176	0.046	0.116	0.054	0.257	0.031
	(1.31)	(0.21)	(0.80)	(0.43)	(0.77)	(0.10)
Three-generation households	0.235	0.078	0.162	0.090	0.096	0.137
	(1.36)	(0.34)	(1.46)	(0.65)	(0.26)	(0.40)
Others	0.200	0.078	0.091	0.115	0.312	0.059
	(1.43)	(0.34)	(0.66)	(0.86)	(1.19)	(0.19)
Age of household head	0.018	0.012	0.012	0.006	0.020	-0.005
	(1.81)	(1.36)	(1.13)	(1.23)	(0.55)	(0.22)
Age of household head, squared	-0.164	-0.125	-0.101	-0.046	-0.185	-0.038
	(1.70)	(1.37)	(0.99)	(0.85)	(0.54)	(0.16)
Education of head of hous	sehold					
Primary school	0.108	0.047	0.133	0.080	0.222	0.072
	(2.45)	(1.22)	(3.54)	(1.50)	(1.79)	(0.68)
Lower secondary	0.169	0.154	0.160	0.211	0.257	0.140
	(2.74)	(3.05)	(3.35)	(2.83)	(2.22)	(1.33)
Upper secondary	0.314	-0.008	0.235	0.217	0.430	0.075
	(3.32)	(0.04)	(2.91)	(2.16)	(2.03)	(0.34)
Short-term technical worker	0.121	0.126	0.233	0.262	-0.509	-0.584
	(1.37)	(1.64)	(2.40)	(3.27)	(2.42)	(1.54)
Long-term technical worker	0.251	0.275	0.285	0.376	0.463	0.213
& professional secondary school	(4.18)	(2.66)	(5.97)	(3.99)	(2.77)	(0.98)
Bachelors and above	1.039	0.251	0.227	0.254		0.663
	(9.33)	(0.72)	(1.80)	(2.72)		(2.86)
Land area per household	(ha)					
Annual land	0.302	0.303	0.158	0.271	0.119	-0.440
Annual land, annual	(5.70)	(2.29)	(3.33)	(2.80)	(0.57)	(1.90)
Annual land, squared	-0.027	-0.058	-0.013	-0.028	0.018	0.206
Development and	(4.36)	(1.08)	(1.81)	(1.01)	(0.27)	(2.17)
Perenniai land	0.281	1.091	0.168	0.304	0.139	0.561
Description of the second second	(1.47)	(2.40)	(2.43)	(2.63)	(0.61)	(2.61)
Perennial land, squared	-0.076	-0.862	-0.010	-0.143	-0.004	-0.082

# Table 6: Expenditure Regressions for Mixed Communes, 1993, 1998 and 2004The Dependent Variable is Log of Real Per Capita Expenditure in January 1998 Prices

	199	93	1998		2004	
	Kinh-Hoa	Minority	Kinh-Hoa	Minority	Kinh-Hoa	Minority
	(1.35)	(2.04)	(2.00)	(2.01)	(0.04)	(2.52)
Forestry land	0.164	3.890	0.032	0.035	-0.002	0.059
	(1.57)	(3.10)	(0.87)	(0.81)	(0.01)	(0.76)
Forestry land, squared	-0.027	-15.46	0.002	-0.003	-0.002	-0.005
	(1.14)	(1.63)	(0.57)	(0.60)	(0.15)	(0.46)
Water-surface land	0.065	0.190	0.204	0.929	-1.706	2.739
	(0.34)	(2.08)	(1.21)	(1.47)	(0.49)	(1.34)
Water-surface land, squared	0.023	-0.056	-0.021	-1.001	7.109	-3.272
	(0.28)	(1.73)	(1.15)	(0.72)	(0.86)	(0.61)
Other land	0.178	0.318	0.001	0.145	0.461	-0.777
	(0.82)	(1.35)	(0.02)	(2.15)	(1.59)	(1.23)
Other land, squared	-0.107	-0.150	0.006	-0.076	-0.119	0.570
	(1.10)	(1.03)	(0.81)	(2.36)	(1.07)	(1.48)
Remittances	0.100	-0.049	0.098	0.066	-0.039	-0.094
	(1.99)	(1.01)	(2.29)	(1.28)	(0.21)	(0.56)
Non-farm self employment	0.36	0.383	0.343	0.322	0.343	0.151
	(3.85)	(2.37)	(4.72)	(3.35)	(1.54)	(0.82)
Constant	7.103	7.067	7.381	7.430	7.149	8.462
	(29.23)	(26.27)	(26.51)	(51.54)	(8.84)	(13.20)
Observations	671	385	666	492	230	223
Number of communes	33	33	37	37	151	151
R-squared	0.32	0.41	0.26	0.42	0.5	0.54

Notes: Commune fixed effects are included in all regressions. The fixed effects sum to zero in each regression. Commune level cluster robust *t*-values in parentheses.

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

In general the estimated parameters have the expected signs and for most of the estimates the order of magnitude is reasonable. Household size has the expected negative effect on log expenditure indicating economies of size, and the proportion of children—both small children (0-6) and older children (7-16)—has the expected negative effect with a larger impact (in absolute terms) of small children than large children in all regressions. Household composition effects on log expenditure are only significantly different (at the 5 percent level) between the Kinh-Hoa and the minorities in the full 1998 sample. For the other regressions, the Chow-tests indicate fairly similar effects of household composition (Table 7). The Chow-tests also clearly indicate that the household type and the age of the household head have the same impacts on log expenditure for Kinh-Hoa and minorities. When tested jointly the total impact of household characteristics (size, composition, type and age of household head) appear to be different for the Kinh-Hoa and the minorities in rural Vietnam, although this difference is only marginally significant, while the effect is probably not different in the mixed communes.

Turning to the returns to endowments, we plot the returns to education in Figures 3 and 4. Figure 3 shows the estimated returns based on the full rural samples while Figure 4 gives the estimated returns for the mixed communes. The return profiles in Figure 3 appear reasonable as the returns are, generally, increasing with the level of education. There is some variation in the profiles with respect to the returns to upper secondary schooling and short-term technical workers but this is not surprising as short-term technical workers have vocational technical education, which can follow either primary school or lower secondary school. Hence, returns to upper secondary schooling and short-term technical training may well be equal. The results for the mixed communes in Figure 4 are less clear. This is probably due to the smaller samples resulting in poorly determined returns for each education

group. Looking at the differences between the returns to Kinh-Hoa and minorities there is no clear pattern, neither in the results based on the full rural samples nor in the mixed commune samples. The similarity of the returns to education across the ethnic groups in Figures 3 and 4 is confirmed by the Chow-tests in Table 7. The test results indicate that returns to education are indeed equal across ethnicity, in particular in 2004. However, it is important to note that the returns to education are relative to the base (no formal education) in each group. This means that, equality of returns does not imply equality of per capita expenditure levels for households with a given level of education across the two groups.

The profiles of the estimated returns to land are shown in Figures 5 and 6, where the return is again measured as the change in log expenditure. The inclusion of squared terms in the regressions only results in mildly decreasing returns to scale for most of the land types. In the full sample the main exceptions are returns to perennial land, returns to water surface land for minorities in 1998 and, in particular, forestry land for the minorities in 1993. In the mixed commune samples some of the point estimates lead to unexpected return profiles, however, as seen from Table 6 all the 'implausible' profiles are related to insignificant point estimates.





*Notes:* The vertical bars are the individual 95 percent confidence bounds. *Source:* Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.





*Notes:* The vertical bars are the individual 95 percent confidence bounds. *Source:* Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.



Figure 5: Estimated Return to Land for Kinh-Hoa and Minorities Rural Vietnam, 1993, 1998 and 2004

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.



### Figure 6: Estimated Return to Land for Kinh-Hoa and Minorities. Mixed Communes, 1993, 1998 and 2004

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

	R	Rural Vietnam		Mixed communes		unes
	1993	1998	2004	1993	1998	2004
Household composition	0.73	0.01	0.08	0.80	0.23	0.98
Household type	0.14	0.32	0.73	0.83	0.63	0.98
Age of household head	0.18	0.90	0.61	0.46	0.90	0.78
Household characteristics, jointly	0.02	0.04	0.08	0.13	0.10	>0.99
Education	0.24	0.19	0.86	0.19	0.37	0.99
Land	0.07	0.00	0.03	0.01	0.01	0.97
Land index	0.89	0.49	0.94	0.21	0.43	0.33
Remittances	0.09	0.91	0.49	0.02	0.63	0.91
Non-farm self employment	0.64	0.57	0.28	0.90	0.86	0.75
All slope parameters	0.00	0.00	0.02	0.00	0.00	>0.99

Table 7: Chow-tests of Equality of Parameters across Kinh-Hoa and Minority
Expenditure Regressions, 1993, 1998 and 2004 ( <i>p</i> -values)

Notes: 'Household composition' is household size, proportion of children 0-6, proportion of children 7-16 and proportion of males above 16. 'Household type' is Parent(s) and one child, Parent(s) and two children, Parent(s) and three or more children, Three generation households and others. 'Age of household head' includes age and age squared. 'Household characteristics, jointly' tests all parameters of the three aforementioned groups jointly. 'Education' consists of the six education indicators. 'Land' consists of the five types of land and their squares. 'Land index' is a weighted average of the different land types with year and group specific weights as explained in the text. *Source:* Own calculations based on the 1993 and 1998 VLSS and the 2004 VHLSS.

In 1993 the minorities had higher returns on both annual and perennial land while Kinh-Hoa had higher returns on forestry land. The return to most land types appears to have decreased for both groups from 1993 to 2004, and in 2004 the differences are small, except for the return to water surface land for which the return is much higher for the minorities compared to the Kinh-Hoa.

In order to get a more comparable measure of the return to land holdings we compute a year and group specific land index. The index is a weighted average of the five different land types using year and group specific averages of the household specific land shares as weights. By this index we aim at measuring a 'typical' return to land holdings for the two groups rather than the return to identical land holdings. As the Kinh-Hoa mostly live in the two deltas and the plains while the minority groups, although scattered, often live in the mountainous areas, identical land compositions are uncommon. If, however, the households are seeking to maximize the return to overall land holdings conditional on location our land index should illustrate the average return to such overall land holdings.<sup>103</sup> The plots in the lower right corners of Figures 5 and 6 show the estimated returns to the land indexes. These returns are less dispersed than the returns to the individual land types illustrating that the average returns to overall land holdings may be equal for the Kinh-Hoa and the minorities.

The pronounced differences in the type specific returns and much smaller differences in returns to 'typical' land holdings are confirmed by the Chow-tests in Table 7. For the full rural samples we reject equality of returns to land in 1998 and 2004, while it cannot be rejected at a 5 percent level in 1993. For the mixed commune samples we reject equality of the returns in 1993 and 1998 while we cannot reject

<sup>&</sup>lt;sup>103</sup> If we consider household production functions in which the different land types are the inputs (generating income/consumption) and assume these production functions are Cobb-Douglas, then the natural choice of weights are the land shares. van de Walle and Guenwardena (2001) also construct land indexes but they use identical land bundles for both groups and the weights are based on the average land holdings of each type. Because of the highly skewed distribution of land types, weights based on the land shares evaluated at the mean land sizes are very different from our weights, which are averages of actual land holding shares.

#### Ethnicity and Poverty Redution

equality in 2004. In contrast, equality of the returns to the land index cannot be rejected in any of the models at conventional levels of significance. This result is in sharp contrast to van de Walle and Gunewardena (2001) who find that the ethnic minorities have higher returns to land than the Kinh-Hoa. The dissimilar results are caused by the different weights used in the construction of the land indexes (see footnote 103).

Also for the land holdings it is important to remember that the estimated returns are relative to group specific base cases (no land). Thus, equality of the returns to typical land holdings does not imply equal expenditure levels only equality of the proportional changes in expenditure relative to the base cases.

The final two regressors are the indicator of remittances and the share of the household engaged in non-farm self-employment. Remittances are insignificant in 7 of the 12 regressions (mainly the minority regressions), and in some regressions the point estimate is negative. The insignificance and the unexpected signs may be an endogeneity, or self selection, problem. Even though we do not include information on the size of remittances there could still be an association between expenditure shocks and remittances. Furthermore, remittances may be a result of deliberate actions of the households to insure against location specific shocks. The large increase over time in the number of households receiving (mainly domestic) remittances indicates that households are in fact actively seeking this kind of insurance. If households self-select into such insurance schemes this will also bias the parameter estimates. Another possible explanation could be that remittances are closely correlated within areas. If this is the case, part of the effect of remittances will be included in the commune fixed effects, and the large standard errors (and possibly even the unexpected signs) are then ordinary multicollinearity problems. Regardless of the potential biases, based on the Chow-tests in Table 7, we find remittances to have fairly similar effects on log expenditure for Kinh-Hoa and the minorities at least in 1998 and 2004.

In contrast to remittances, non-farm self-employment has a sizable impact on expenditure. In the full sample regressions the parameter estimates are around 0.3 indicating that, say, in a household of 4 per capita expenditure is about 8 percent higher if one member is engaged in non-farm self-employment compared to a family in which no members are engaged in this kind of income generation. Comparing the returns across the two groups we cannot reject that the effect of non-farm self-employment on log expenditure is equal in any of the sample years.

Summarizing the regression results we find a somewhat surprising similarity between the Kinh-Hoa and the minority expenditure regressions within each year. The main difference appear to be the return to the different land types and this may in part be explained by differences in land-use in different locations although the differences also persist in the mixed commune samples in 1993 and 1998. Despite the probable equality of some of the returns the overall Chow-tests in Table 7 indicate that the reduced form expenditure models for Kinh-Hoa and the minorities are different, except in the mixed communes in 2004 where we find strong indications of equal slope parameters. Based on these results we move on to look at decompositions of the average expenditure gaps using the Blinder-Oaxaca decomposition technique.

### 4.3. Decomposition of the Expenditure Gaps

The Blinder-Oaxaca decomposition (Blinder 1973, Oaxaca 1973) is a well known method often used in labor economics to test for labor market discrimination. van de Walle and Gunewardena (2001) used the decomposition to analyze the extent of discrimination of ethnic minorities in Vietnam based on the 1993 survey and Baulch *et al.* (2004) updated the analysis to cover the 1998 survey using a slightly different regression specification. Even though the decomposition technique is well known (see *e.g.*, Greene 2003) there are some special problems when applying it to fixed effects models. Therefore, we briefly describe the decomposition technique before the results are presented and discussed.

The basic idea in the Binder-Oaxaca decomposition is to split the average log expenditure gap between the Kinh-Hoa majority and the minorities into two parts, one being the gap caused by differences in endowments (the regressors) and the other being the gap caused by differences in returns (the regression parameters). As both of these differences are actually vectors of differences a method of aggregating the individual contributions must be chosen. Based on Oaxaca (1973) the following decomposition is made

$$\overline{\ln(E_{hc1})} - \overline{\ln(E_{hc2})} = \overline{x}_{hc1}\hat{\beta}_1 - \overline{x}_{hc2}\hat{\beta}_2$$

$$= (\overline{x}_{hc1} - \overline{x}_{hc2})\tilde{\beta} + \overline{x}_{hc1}(\hat{\beta}_1 - \tilde{\beta}) + \overline{x}_{hc2}(\tilde{\beta} - \hat{\beta}_2)$$

$$= [characteristics] + [majority advantage] + [minority disadvantage]$$
(2)

In equation subscript '1' denotes the Kinh-Hoa regression while subscript '2' denotes the minority regression and the bars specify averages. The first equality in equation is given from the fact that the average of the residuals is zero, by construction, in OLS (and hence fixed effects) regressions,  $\hat{\vec{E}}_{hcg} = 0$ , g = 1, 2, and, further, using the identifying restriction that the fixed commune effects sum to zero in each group

$$\sum_{c=1}^{+g} \eta_{cg} = 0 \text{ for each } g = 1,2$$
(3)

The second equality in is the Blinder-Oaxaca decomposition in which  $\hat{\beta}$  is an estimate of the nondiscriminatory return structure. Inserting the nondiscriminatory return structure into the first line in and rearranging gives the second line whereby the log expenditure gap is decomposed into an estimate of the endowment differential, an estimate of the Kinh-Hoa return advantage and an estimate of the minority return disadvantage. The sum of the last two terms is often called the unexplained part of the gap and related to discrimination because it is the gap in expenditure that is 'caused' by differences in returns to endowments.

Following Oaxaca and Ransom (1988, 1994) the estimated nondiscriminatory return parameters can be expressed as a matrix weighted average of the estimated Kinh-Hoa and minority parameters

$$\tilde{\beta} = W\hat{\beta}_1 + (I - W)\hat{\beta}_2,\tag{4}$$

where *W* is a *k* x *k* weighting matrix and *I* is the *k* x *k* identity matrix. In most of the early studies applying the decomposition the nondiscriminatory return parameters were assumed to be equal to either the majority parameters or the minority parameters as these are the most natural bounds for the nondiscriminatory structure and this is also the assumption made in both van de Walle and Gunewadena (2001) and Baluch *et al.* (2004). In terms of the weighting matrix this is equivalent to setting W = I (majority structure) and W = 0 (minority structure).<sup>104</sup>

Heitmüller (2005) has noted an important identification problem when the Blinder-Oaxaca decomposition is applied to fixed effects models as we do below. To illustrate the problem we first partition the regressors in equation into  $x_{hcg} = [1, x_{hcg}^{l}]$  and, accordingly, the parameter vector into.  $\beta'_{g} = [\beta_{g}^{0}, \beta_{g}^{l'}]$  Moreover, we consider a set of commune specific regressors, which are often included in reduced form expenditure regressions to capture important location effects. The most common commune characteristics are infrastructure (quality of roads, public transportation, and distance to markets, government offices, schools, hospitals and so forth). Other natural candidates are related to the quality of land. As we are using fixed effects regressions the specific choice of commune characteristics is not important. Hence, we simply specify a 1 x *l* vector,  $z_{3}$ , and an associated reduced form *l* x 1 parameter vector  $\theta_{8}$ . Inserting the commune characteristics into the expenditure regression we obtain

 $C_{-}$ 

<sup>&</sup>lt;sup>104</sup> Several other weighting matrices have been proposed in the literature. See Oaxaca and Ransom (1994) for a discussion.

Ethnicity and Poverty Redution

$$\ln(E_{hcg}) = x_{hcg}^1 \beta_g^1 + \gamma_g + z_{cg} \theta_g + \alpha_{cg} + \varepsilon_{hcg},$$
(5)

where  $\gamma_3$  is the new intercept in the regression and  $\alpha_{c_3}$  are the latent commune specific factors. Using the analogue identifying restriction for the commune specific effects as above

$$\sum_{c=1}^{C_g} \alpha_{cg} = 0, \quad \text{for each } g = 1,2 \tag{6}$$

we can relate the fixed effects in equations (1) and (5) and hence the intercepts

$$\eta_{cg} = (z_{cg} - \overline{z}_{cg})\theta_g + \alpha_{cg} \tag{7}$$

where  $\overline{z}_{cg}$  is the average of the commune regressors over all communes and

$$\beta_g^0 = \gamma_g + \overline{z}_{cg} \theta_g. \tag{8}$$

Finally, substituting the expression for the intercept into the decomposition in (2) and rearranging we obtain an augmented decomposition that includes a decomposition of the commune specific regressors

$$\overline{\ln(E_{hc1})} - \overline{\ln(E_{hc2})} = (\overline{x}_{hc1}^1 - \overline{x}_{hc2}^1)\widetilde{\beta}^1 + \overline{x}_{hc1}^1(\widehat{\beta}_1^1 - \widetilde{\beta}^1) + (\widehat{\gamma}_1 - \widetilde{\gamma}) + \overline{x}_{hc2}^1(\widetilde{\beta}^1 - \widehat{\beta}_2^1) + (\widetilde{\gamma} - \widehat{\gamma}_2)$$
$$+ (\overline{z}_{c1} - \overline{z}_{c2})\widetilde{\theta} + \overline{z}_{c1}(\widehat{\theta}_1 - \widetilde{\theta}) + \overline{z}_{c2}(\widetilde{\theta} - \widehat{\theta}_2)$$
(9)

where we again use a bar to indicate averages and a tilde to indicate the estimated nondiscriminatory return parameters.

The decomposition in (9) shows that the original decomposition in must be interpreted with caution in the presence of commune fixed effects. The constant term in, which is part of the difference in returns, is confounded as it includes the difference between the intercepts at the household level  $(\overline{z}_{ch1} - \overline{z}_{ch1})\tilde{\theta}$ , and the commune effects; both commune endowments,  $(\hat{\gamma}_1 - \hat{\gamma}_2)$  and differences in returns to commune endowments  $\overline{z}_{cl}(\hat{\theta}_1 - \tilde{\theta})_+ \overline{z}_{c2}(\tilde{\theta} - \hat{\theta}_2)$ . This means that when the decomposition in is interpreted in a strict way, the results are biased. But, if we instead interpret the decomposition as splitting the expenditure gap into a part explained by *household specific endowments* and the remainder part, being household specific returns *and* location effects (both endowment and differences in returns), then the decomposition is 'unbiased'. Needless to say, it is important to seek to isolate the commune endowment effect in order to fully understand the factors behind the expenditure gap but this is outside the scope of this chapter. Yet, as seen from the decomposition in, when we consider the mixed commune samples the differences in commune specific endowments,  $(\overline{z}_{ch1} - \overline{z}_{ch2})\tilde{\theta}$ , is zero by construction. Hence, for these samples the commune level endowment effect is removed whereby the normal interpretation of the decomposition makes sense as all terms in the intercept are related to differences in returns, some at the household level and some at the commune level.

Table 8 reports the results of the decomposition using first the estimated parameters from the Kinh-Hoa regression and second the estimated parameters from the minority regression as the nondiscriminatory return estimates. The table also includes both the full rural samples (top part) and the mixed commune samples (bottom part).

	1	993	1998		2004	
	Estimates	Percent	Estimates	Percent	Estimates	Percent
Rural Vietnam						
Difference	0.355	100	0.466	100	0.558	100
	(0.031)		(0.028)		(0.013)	
Kinh-Hoa as reference ( $W =$	I)					
Household Characteristics	0.080	23	0.108	23	0.102	18
	(0.025)	[09;37]	(0.032)	[10;37]	(0.015)	[13;24]
Household returns and	0.275	77	0.358	77	0.456	82
location effects	(0.021)	[63;91]	(0.024)	[63;90]	(0.014)	[76;87]
Minorities as reference ( $W =$	0)	• • •		• • •	. ,	• • •
Household characteristics	-0.188	-53	0.052	11	0.121	22
	(0.235)	[-182;77]	(0.121)	[-40;62]	(0.025)	[13;30]
Household returns and	0.543	153	0.414	89	0.437	78
location effects	(0.234)	[23;282]	(0.118)	[38;140]	(0.023)	[70;87]
Mixed Communes		• • •		• • •	. ,	• • •
Difference	0.323	100	0.342	100	0.247	100
	(0.044)		(0.036)		(0.068)	
Kinh-Hoa as reference ( $W =$	I) $($		. ,		. ,	
Household characteristics	0.050	16	0.088	26	0.125	51
	(0.039)	[-8;39]	(0.038)	[4;48]	(0.076)	[-18;120]
Household returns and	0.273	84	0.255	74	0.122	49
location returns	(0.027)	[61;102]	(0.028)	[52;96]	(0.087)	[-20;118]
Minorities as reference ( $W =$	0)					
Household characteristics	-0.756	-234	-0.171	-50	0.134	54
	(0.984)	[-834;365]	(0.492)	[-332;232]	(0.085)	[-22;113]
Household returns and	1.079	334	0.513	150	0.113	46
location returns	(0.986)	[-265;934]	(0.492)	[-132;432]	(0.081)	[-13;122]

# Table 8: Decomposition of the Log Per Capita Expenditure Difference between Kinh-Hoa and Minorities, 1993, 1998 and 2004

Note: Standard errors in parentheses. Approximate 95 percent confidence bounds in brackets.

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

Looking first at the results for the full rural samples it is evident that the expenditure gap has been increasing from 1993 to 2004, reflecting the results in Table 2. In 1993 average Kinh-Hoa expenditure was about 43 percent higher than the average minority expenditure. The gap increased to 59 percent in 1998 and 75 percent in 2004. When the Kinh-Hoa return parameters are used as estimates of the nondiscriminatory return structure the relative importance of household endowments differences is surprisingly constant over time as the endowment gap constitutes about 20 percent in all three years. When the minorities' return parameters are used as nondiscriminatory reference the household endowment differences are even less important in 1993 and 1998, but this decomposition leads to poorly determined estimates. In 2004 the choice of reference return structure does significantly change the decomposition results. Hence, the full rural sample results strongly indicate that policies aimed at decreasing the family size, increasing the educational level of the household head and improving the outside job options for the ethnic minorities will only have a limited effect on the average expenditure gap and, hence, on the gap in the poverty incidences.

Concluding that the gap is mainly due to differences in the expenditure generation models leading to radically different returns to human capital and land (which may be caused by discrimination

against the minorities) is, however, wrong because the main element of the 80 percent of the gap that is attributed to differences in returns and location may well be the location effect. One indication of the importance of location is given by the results obtained for the samples of mixed communes.

In the mixed commune samples the expenditure gaps are generally smaller than the full sample gaps and, furthermore, the gap is fairly constant, around 40 percent in 1993 and 1998, subsequently falling to 28 percent in 2004. Statistically, we cannot exclude that the gap has been constant, however, an increase in the gap, as seen for the full sample, is very unlikely. In 1993 and 1998, the relative contribution from household characteristics supports the full sample results, being close to 20 percent. Hence, eliminating the average location difference has an impact on the size of the gap but not the relative distribution in 1993 and 1998.

The picture is different in 2004 as expected from the results in Tables 6 and 7. In 2004 the point estimate of the relative importance of household endowments is about 50 percent regardless of the chosen reference return structure. But, the effects are not well determined, and we cannot reject the null hypothesis that the gap is solely caused by differences in household endowment differences. On the other hand we also cannot reject the null hypothesis that the gap is only caused by differences in returns. But the latter hypothesis is not supported by the regression results in Tables 6 and 7.

The results for the mixed communes do not imply that past adverse (discriminatory) attitudes towards the minority groups have decreased or disappeared as the differences in household endowments, in particular education and participation in off farm employment, shown in Table 4, may be a result of external constraints such as limited access to credit. However, the results do show that policies aimed at increasing the educational level of the household head and improving the outside job options for the ethnic minorities in the mixed communes would probably narrow the expenditure gap considerably.

	1993		1998		2004	
Education	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Kinh-Hoa Majority						
Less than primary	0.10	0.25	0.13	0.28	0.16	0.30
Primary school	0.14	0.27	0.14	0.28	0.20	0.33
Lower secondary	0.13	0.29	0.17	0.31	0.19	0.32
Upper secondary	0.15	0.30	0.17	0.31	0.25	0.36
Short-term technical worker	0.17	0.30	0.15	0.28	0.35	0.40
Long-term technical worker						
& professional secondary school	0.14	0.28	0.15	0.28	0.19	0.32
Bachelors and above	0.18	0.35	0.10	0.27	0.11	0.25
Total	0.13	0.27	0.16	0.29	0.20	0.32
F-test (p-value)	0.008		0.054		< 0.000	
Minorities						
Less than primary	0.03	0.14	0.03	0.14	0.03	0.13
Primary school	0.02	0.10	0.03	0.12	0.06	0.21
Lower secondary	0.01	0.07	0.03	0.13	0.07	0.21
Upper secondary	-	-	0.03	0.10	0.05	0.16
Short-term technical worker	0.04	0.13	0.00	0.00	0.04	0.14
Long-term technical worker						
& professional secondary school	0.01	0.05	0.01	0.06	0.10	0.24
Bachelors and above	0.13	0.25	0.06	0.11	0.24	0.23
Total	0.02	0.12	0.03	0.13	0.05	0.18
F-test (p-value)	0.537		0.894		< 0.000	

#### Table 9: Non-Farm Self-Employment by the Level of Education of the Household Head

Source: Own calculations based on 1993 and 1998 VLSS and 2004 VHLSS.

Unfortunately, initiatives aimed at improving the outside job options may be difficult to implement in the short run as education and participation in non-farm self-employment are interdependent. In Table 9 we report the average household participation in non-farm self-employment by the educational level of the household head and the associated ANOVA *F*-tests. For the Kinh-Hoa majority there is a clear, statistically significant, relationship between education and non-farm work in all three survey years showing that this type of work is most prevalent among the better educated households, in particular households in which the household head has completed secondary school and/or technical training. For the minority people non-farm work is rare in general, as noted in Table 3, and the results in Table 9 indicate that the increase in this type of work from 1998 to 2004 is, in relative terms, mainly due to increases in non-farm self-employment for the minority people in the rural areas may not be pro-poor policies and they may not be efficient if the educational level of the household head is a binding constraint.

The combination of the results for the full rural sample and the mixed communes in 2004 indicates strong location effects on the average expenditure gap. Although we cannot rule out the possibility that minority people in the mixed communes could be 'behaviorally' closer to the Kinh-Hoa than to the minority people in the pure minority communes whereby the location effect is not simply a matter of infrastructure and land quality it still points to a strong geographical element in the expenditure gap. In addition, the Chow-tests in Table 7 indicate that the expenditure model for the minorities is not that different from the model for the Kinh-Hoa in 2004 supporting the view that location is an important obstacle to increased well being for the minority people. Hence, increased focus on national programs targeted towards the pure minority communes, such as the new program launched in January 2006 that covers all the mountainous and ethnic minority areas, appears highly warranted.

### 5. Summary and Concluding Remarks

Large differences in poverty rates and other socio-economic measures between the major ethnic group (the Kinh people often grouped with the Hoa people) and the other 52 ethnic minority groups living in Vietnam is a well established fact, which has long been recognized by research observers and by the Government of Vietnam. Quantitative analyses from the late 1990s and early 2000s, based on the 1993 and 1998 VLSSs, point to three interlinked sources of the large, and increasing, discrepancy in majority and minority well-being. One factor is differences in household characteristics and endowments such as family size and composition, education, agricultural land holdings and non-farm employment. The second, equally important, factor is a marked dissimilarity in the models generating the household incomes leading to differences in returns to productive endowments such as education and land. The third factor, which is influencing the first two, is differences in location.

In this chapter we contribute to the quantitative work in two ways. First of all we update the quantitative analyses to include the latest household survey information (VHLSS 2004) and, second, we seek to illustrate changes over time in the relative importance of household endowments, the returns to household endowments and household location by using directly comparable empirical models for per capita consumption expenditure and comparable samples in an econometric analysis of the expenditure gaps between majority and minority households.

Using household level data from 1993, 1998 and 2004 we find the decline in poverty among the ethnic minorities to be markedly larger in the second half of the period compared to the first half. But it is still not as large as the decline in poverty among the majority people. Moreover, looking at average per capita consumption expenditure, which is the measure used in estimating the official GSO poverty rates, we find divergence as the expenditure gap between the majority and minority people has increased both from 1993 to 1998 and further from 1998 to 2004. Hence, although all ethnic groups

have experienced increasing well being over the last 10-15 years the majority people have done far better than the minority people.

Household characteristics differ between the majority and minority people. Minority households typically have more members than majority households. Although the average household size has been decreasing over time for both groups the difference in size still persist in 2004 and in certain aspects the differences have increased over time. In 2004, minority households have more children and there is a higher frequency of three generation households compared to majority households.

Another well-known and much discussed difference between the majority and the minority people is the level of formal education. The educational level of both groups has increased significantly over the period. In particular, there has been an increase in the proportion of household members who have completed primary school or above. But the support for education among ethnic minorities and in extremely difficult areas is mainly targeted towards the children and the young household members. Hence, there is a surprisingly constant fraction, about 50 percent, of minority households in which the household head has no formal education. In contrast, there has been a large drop in that fraction in the majority households.

The distribution of land within and across ethnic groups is more difficult to interpret. On average the minority households have larger areas of annual and perennial land than the majority households and for both groups the average size of the land areas are increasing over time. However, within each group the distribution of land holdings is highly skewed and it is well known that the quality of land varies greatly across Vietnam with more productive land in the deltas and on the plains compared to the mountainous areas. The large within-group variation and the variation in land quality imply that we cannot infer that larger average land holdings are advantageous for the minority people as such.

It is easier to understand the importance of non-farm employment and we find a much larger frequency of non-farm employment among the majority households than the minority households. Moreover, the difference is increasing over time.

Altogether our descriptive analysis points towards a pattern of divergence in household characteristics and endowments, except possibly in terms of child education, which is important but not necessarily an income generating factor in the short run.

We include the above mentioned household characteristics and endowments as regressors in an econometric analysis of the gap in per capita real expenditure in which we condition on location effects by including commune fixed effects. We also seek to 'remove' the location effects by analyzing sub-samples of communes in which the majority and minority are both present (mixed communes). The regression results show somewhat surprising similarities in the expenditure models, which is in contrast to earlier analyses of the expenditure gap by van de Walle and Gunewardena (2001) and Baulch *et al.* (2004). In particular, we do not find significant differences in the returns to education or in the returns to non-farm employment.

The regression models for the Kinh-Hoa majority and the minority groups are used to decompose the average per capita real expenditure gap into factors related to household composition and endowments (household characteristics) and factors related to returns to the household characteristics. We explain in detail that the results of the decompositions should be interpreted with care because of the inclusion of commune fixed effects. The problem is that the effect of different locations is included in the difference between the intercept terms, which is subsequently interpreted as a part of the differences in returns in the classical Oaxaca decomposition. However, location is part of the household endowment and, as such, not all of it should be attributed to differences in returns. Hence, the importance of returns is overestimated while the importance of endowments is underestimated if the decomposition is interpreted in a strict sense. Nevertheless, the decomposition results are still useful because the share of the gap caused by differences in returns can simply be reinterpreted as the sum of return differences and location differences.
The decomposition analysis supports the descriptive results as the gap in per capita real expenditure is increasing over time. Moreover, the relative contribution of household characteristics and returns plus location effects is surprisingly constant as household characteristics only account for about 20 percent of the expenditure gap in all three years. When the location effect is removed, by analyzing the sub-sample of mixed communes, the results change somewhat as the gap is decreasing from 1998 to 2004. But it is still significant in 2004. The most interesting result of the mixed commune analysis is that in 2004 the expenditure gap is probably only caused by differences in household endowments.

Overall, the decompositions point to a strong impact of location on the average expenditure gap supporting the view that location is an important obstacle to increased well being for the minority people. Yet, a pure 'ethnicity factor' cannot be ruled out since both the descriptive analysis and the econometric results clearly illustrate differences in household endowments in communes where Kinh-Hoa and minority people live together.

The strong effect of location on ethnic well being calls for continued and increased government effort in extremely difficult ethnic and mountainous areas. Targeted programs such as Program 135, which was initiated in 1998, has been found to be well targeted but the coverage of the program is probably too limited. Hence, a scaling up of the most successful parts of that program is probably a good idea.

Another aspect of the location effect is that it is important to 'localize' government support programs by taking the local environment into account in the design of specific programs and this calls for in-depth analyses of income and poverty at the regional level and possibly even at lower geographical levels. One such regional analysis is the study of the Northern Uplands of Vietnam by Minot *et al.* (2004). An interesting conclusion in that study is that a "rural development strategy [in the Northern Uplands] must focus on ways to increase the labor and land productivity of crop production, including yield increases, diversification to high-value crops, and other means of increasing the economic returns to crop production" (Minot *et al.* 2004, p. 182). This conclusion squares well with the results of the present analysis because increasing the productivity of agricultural land would be increasing the return to the only assets that seems to be in relative abundance among the ethnic minority people. Therefore, such a development strategy would be likely to benefit the ethnic minority groups in the remote areas.

#### References

- Baulch, B., D. Haughton, J. Haughton, and T. T. K. Chuyen. 2004. "Ethnic Minority Development in Vietnam: A Socioeconomic Perspective". Chapter 8 in P. Glewwe, N. Agrawal and D. Dollar (Eds.), Economic Growth, Poverty, and Household Welfare in Vietnam, Washington DC: The World Bank.
- Blinder, A. S. 1973. "Wage discrimination: reduced form and structural estimates", *Journal of Human Resources*, 8:436-455.
- Bourguinon, F. and L. A. P. da Silva. 2003. The Impact of Economic Policies on Poverty and Income Distribution. Evaluation Techniques and Tools. New York: Oxford University Press.
- Brandt, L. et al. 2005. "Summary report: land Access, land markets and their distributive implications in rural Vietnam". Hanoi. Processed.
- CAF [Centre for Analysis and Forecasting]. 2006. "Poverty and Poverty reduction in Vietnam 1993-2004: An Update". Hanoi. Processed.
- Greene, W. H. 2003. Econometric Analysis. Fifth Edition. New Jersey: Pearson Education, Inc.
- GSO [General Statistics Office]. 2001. Population and Housing Census Vietnam 1999 Completed Census Results. Statistical Publishing House, Hanoi.

- Heitmüller, A. 2005. "A note on decompositions in fixed effects models in the presence of timeinvariant characteristics", IZA Discussion Paper No. 1886. Institute for the Study of Labor, Bonn.
- Hoang T. T. H. 2006. "Performance of non-farm household enterprises (NFHEs) in VHLSS 2004 and VHLSS 2002", Background Report, The World Bank, Hanoi. Processed.
- Minot, M., M. Epprecht, D. Roland-Holst, Tran T. T. A and Le Q. T. 2004. "Income diversification and poverty in the Northern Uplands of Vietnam", International Food Policy Research Institute and Japan Bank for International Cooperation. Agricultural Publishing House in Hanoi.
- MOLISA and UNDP. 2004. "Taking stock, planning ahead: Evaluation of the national targeted programme on hunger eradication and poverty reduction and programme 135", Hanoi.
- Nguyen, H. T. M. 2006. "Ethnic inequality in Vietnam: evidence from micro-level analysis", Research Essay IDEC 8011, Asia Pacific School of Economics and Government, the Australian National University.
- Oaxaca, R. 1973. "Male-female wage differentials in urban labour markets", *International Economic Review*, 14:693-709.
- Oaxaca, R. and M. Ransom (1988) "Searching for the effect of unionism on the wages of union and nonunion workers", *Journal of Labor Research*, 9:139-148.
- Oaxaca, R. and M. Ransom (1994) "On discrimination and the decomposition of wage differentials", Journal of Econometrics, 61:5-21.
- SRV [Socialist Republic of Vietnam]. 2001. *National Target Programme on HEPR and Employment for the Period 2001-2005*, Socialist Republic of Viet Nam, Ministry of Labour, Invalids and Social Affairs, Hanoi.
- van de Walle, D. & D. Gunewardena. 2001. "Sources of ethnic inequality in Viet Nam", Journal of Development Economics, 65:177-207.
- Vijverberg, W. P. M. et al (2006) "Non-Farm Household Enterprises in Vietnam A Research Project using Data from VHLSS 2004, VHLSS 2002 and AHBS 2003", The World Bank, August 2003.
- World Bank (2003) Vietnam Development report 2004: Poverty. Hanoi.

## Appendix 1. Some Legislative Documents for Ethnic Minorities and Mountainous Areas

- Decision 166/2001/QD-UBDTMN on 5 September 2001 promulgating the criteria of special difficult ethnic minority households.
- Circular 912/2001/TTLT/UBDTMN/BTC/BKHDT on 16 November 2001 guiding the implementation of supporting policies for special difficult ethnic minority households.
- Decision 194/2002/QD-TTg 21 December 2001 on adjusting the scholarship and social benefit for ethnic minority pupils, students studying in public schools regulated at Decision 1121/1997/QD-TTg 23 December 1997 of the Prime Minister.
- Decision 1637/QD-TTg 31 December 2001 on granting and delivering newspapers and magazines for the ethnic minorities and mountainous areas.
- Decree 02/2002/ND-CP 3 January 2002 on adjusting, supplementing the Decree 20/1998/NĐ-CP 31 March 1998 of the Government on the development of trade in mountainous, island and ethnic minorities areas.
- Decision 63/2002/QD-TCBD 18 January on assigning the implementation of Decision 1637/QD-TTg 31 December 2001 of the Prime Minister on granting and delivering newspapers and magazines for the ethnic minorities and mountainous areas.
- Circular 13/2002/TTLT/BGDDT-BTC 28 March 2002 on guiding the implementation of Decision 194/2002/QD-TTg 21 December 2001 of the Prime Minister on adjusting the scholarship and social benefit for ethnic minority pupils, students studying in public schools regulated at Decision 1121/1997/QD-TTg 23 December 1997 of the Prime Minister.
- Circular 04/2002/TT-NHNN 3 July 2002 of the Governor of SBV guiding the implementation of interest reduction of State-owned Commercial Banks for businessman in Area II, III of the mountainous, ethnic minorities regulated at Decree 02/2002<sup>ND</sup>-CP 3 January of the Government
- Decision 120/2002/QD-TTg 18 September 2002 on supplementing and changing the name of communes in the socio-economic development program of special difficult communes, ethnic minorities, mountainous, border and remote areas.
- Official letter 1433/VPCP-DPI on 28 March 2003 regulating some policies for ethnic minority areas.
- Circular 56/2003/TT-BNN on 9 April 2003 on guiding the implementation of projects in the Hunger Eradication and Poverty Alleviation and Job Creation directed by Ministry of Agricultural and Rural Development.
- Decision 391/QD-NHNN on 28 April 2003 adjusting Point 3 of the Circular 04/2002/TT-NHNN 3 July 2002 of the Governor of SBV guiding the implementation of interest reduction of Stateowned Commercial Banks for businessman in Area II, III of the mountainous, ethnic minorities regulated at Decree 02/2002<sup>ND</sup>-CP 3 January of the Government
- Decree 51/2003/ND-CP on 16 May 2003 regulating the functions, tasks, powers and organizational structure of The Committee for Ethnic Minority Affairs (CEMA).
- Decision 122/2003/QD-TTg 12 June 2003 on the Government Action Program to implement the Resolution of the 7<sup>th</sup> Communist Party Congress on ethnic issues.
- Decision 124/2003/QD-TTg 17 June 2003 approving the project for protection and development of ethnic minorities cultures in Vietnam.
- Direction 16/2003/CT-TTg on 18 June 2003 improving the management, direction and implementation of socio-economic development program of special difficult communes, ethnic minorities, border and remote areas (Program 135).
- Decision 134/2004/QD-TTg on 20 July regulating some policies on supporting cultivated land, housing land and water for ethnic and disadvantaged minorities

#### Ethnicity and Poverty Redution

- Official letter 2691/BNN-HTX on 8 November 2004 guiding the implementation of 2005 plan on projects of the Hunger Eradication and Poverty Reduction Program and Program 135.
- Direction 38/CT-TTg on 9 November 2004 stepping up the training and fostering of ethnic minority languages for officials and public servants working in ethnic minority and mountainous areas.
- Circular 819/2004/ TTLT/UBDT-KHDT-TC-XD-NNPHNT on 10 November 2004 guiding the implementation of Decision 134/2004/QD-TTg on 20 July regulating some policies on supporting cultivated land, housing land and water for ethnic and disadvantaged minorities.
- Decision 03/2005/QD-BNN on 7 January 2005 promulgating the regulation on exploiting woods to support for poor and difficult ethnic minorities regulated at Decision 134/2004/QD-TTg on 20 July 2004.
- Decision 15/2005/QD-TTg on 19 January 2005 supplementing special difficult communes on the socio-economic development program of special difficult, ethnic minorities, border and remote areas.
- Decision No. 170/2005/QD-TTg promulgating the poverty line for the 2006-2010 period.
- Decision 393/2005/QD-UBDT on 29 August 2005 promulgating criteria classifying ethnic minorities and mountainous areas based on their development level.
- Decision 267/2005/QD-TTg on 31 October 2005 regulating vocational training for ethnic minority pupils at boarding schools.
- Decision 304/QD-TTg on 23 November 2005 piloting the handing over and protecting the forest for residential ethnic minority households and communities in the Central Highlands.
- Decision 07/2006/QD-TTg 10 January 2006 approving the socio-economic development program for special difficult, ethnic minorities and mountainous areas in 2006-2010.
- Decision 82/2006/QD-TTg 14 April 2006 adjusting the policy scholarship for ethnic minority pupils, students studying in boarding high schools and probationary universities regulated at Decision 194/2001/QD-TTg of the Prime Minister on 21 December 2001.

orthern mountainous area Central Highlands		Khmer	Othe	Other minorities		
Tày	Gia rai	Khơ me	C $\sigma$ ho	Со		
Thái	Ê đê		Chăm	Tà oi		
Mường	Bana		Hrê	Cho ro		
Nùng	$X \sigma$ đăng		Mnông	Xinh mun		
H'Mông	Giẻ triên		Raglai	Hà Nhì		
Dao	Mạ		Xtiêng	Lào		
Sán chay	Chu ru		Bru Vân K	La Chí		
Sán Dìu			Thổ	La Ha		
			Giáy	Lự		
			Cơ tu	Cống		
			Khơ mú			

 Table A: The Geographical Grouping of the Ethnic Minorities

### Appendix 2. Geographical Grouping of the Ethnic Minorities

# 6

### Changes in Household's Poverty Status in Rural Vietnam 2002-2004

#### Vu Hoang Dat, Truong Thi Thu Trang, Nguyen Viet Duc and Henrik Hansen

#### 1. Introduction

Vietnam has achieved remarkable success in reducing poverty in the period 1993-2004. The consumption-based poverty rate, which is estimated from nationally representative household surveys, has come down from 58.1 percent in 1993 to 19.5 percent in 2004, a decline of almost 40 percentage points over the eleven years for which it has been measured. In absolute terms the decline in the poverty rate can be translated to the result that some twenty-four million people became non-poor at some point during the years from 1993 to 2004. Approximately half of them moved out of poverty during the first five years, 1993-1998, while the other half crossed the poverty line in the second sub-period, 1998-2004. Digging a little deeper into the poverty figures, it is well known that the decline in poverty varies substantially across geographical regions, urban and rural areas, and ethnic groups (Table 1) and this variation has been the subject of many recent analyses (see e.g., World Bank 2003 and VASS 2006).

Year	1993	1998	2002	2004
Vietnam	58.1	37.4	28.9	19.5
Urban	25.1	9.2	6.6	3.6
Rural	66.4	45.5	35.6	25.0
Kinh & Hoa	53.9	31.1	23.1	13.5
Other ethnic groups	86.4	75.2	69.3	60.7
Northern Uplands	81.5	64.2	43.9	35.4
Read River Delta	62.7	29.3	22.4	12.1
North Central Coast	74.5	48.1	43.9	31.9
South Central Coast	47.2	34.5	25.2	19.0
Central Highlands	70.0	52.4	51.8	33.1
South East	37.0	12.2	10.6	5.4
Mekong Delta	47.1	36.9	23.4	15.9

Source: VASS (2006)

Yet, though informative, the disaggregated poverty rates in Table 1 do not tell the whole story about poverty in Vietnam as they only show the net changes in the poverty rates. In particular the poverty rates do not give information about the gross flows in and out of poverty. Hence, they do not reveal important information about the structure of the Vietnamese poverty such as the share of chronic poor or the characteristics of chronic vs. transient poor. Such disaggregated poverty analyses can provide important inputs into the design and re-design of pro-poor policies to take into account the significant level of heterogeneity among the poor and also factors that help reducing the vulnerability of the non-poor.

Given the significant policy importance, analyses of poverty dynamics have attracted considerable attention from researchers over the last decade or so. On the data side, availability of panel data, which is a prerequisite for the analysis, has clearly improved. As a result, a growing number of studies analyzing poverty dynamics and its determinants in developing and transition countries have appeared in recent years. Specifically, poverty dynamics has been analyzed for China (Jalan and Ravallion, 1998; McCulloch and Calandrino, 2003), Ethiopia (Dearcon and Krishnan, 2000; Kedir and McKay, 2005), Madagascar (Herrera and Roubaud, 2005), Pakistan (Baulch and McCulloch, 1998), Peru (Glewwe and Hall, 1995; Herrera, 2001), Poland (Okrasa, 1999), South Africa (Carter and May, 2001), Uganda (Lawson *et al.*, 2005) and Venezuela (Freije, 2000). While this list of studies on poverty dynamics in developing and transition countries is clearly incomplete it nonetheless provides a useful and rich source of information on relevant knowledge accumulated to date.

In Vietnam, poverty dynamics during the period 1993-1998 have been studied by Glewwe Gragnolati and Zaman (2002), Justino and Litchfield (2003), Litchfield and Justino (2004) and Nimii Vasudeva-Dutta and Winters (2003, 2004). The studies all use the panel data set drawn from the household surveys VLSS 1993 and VLSS 1998. Thus the information that can be drawn from these studies may not be relevant for current poverty reduction policies considering the rapid development of the Vietnamese economy and the differences in poverty levels in the 1990s and the 2000s. Therefore, in this chapter, we update the analysis of the poverty dynamics and its determinants in rural Vietnam to the period 2002-2004 for which the second set of panel households are available from the Vietnam Household Living Standard Surveys (VHLSS) 2002 and 2004. The aim is to establish an 'extended' poverty profile that includes the gross flows in and out of poverty in the period. As such, the study will hopefully contribute to further understanding of the determinants of the more recent poverty dynamics in Vietnam and, hence, provide a basis for policies that can maintain the pro-poor pattern of economic growth.

The chapter is organized as follows. In Section 2 we define and describe the poverty transitions by way of cross tabulations. In section 3 we formulate and estimate an econometric model for the poverty transitions. The model is inspired by, and largely follows, the modeling strategy formulated in the studies, mentioned above, of the poverty dynamics in Vietnam during the 1990s. Section 4 offers some conclusions as well as suggestions for further research.

#### 2. Poverty Transitions from 2002 to 2004

In this section we describe the recent poverty transitions in rural Vietnam. The reason for focusing on the rural parts of Vietnam and, hence, excluding the urban areas from the analysis can be seen from Table 1. While poverty was wide spread in the early 1990s the substantial decline during the 1990s and early 2000s has made poverty a predominantly rural problem. The urban poverty rate was 6.6 percent in 2002 dropping further to 3.6 percent in 2004. In contrast, the rural poverty rate was 35.6 percent in 2002, about five times the urban rate, falling to 25 percent in 2004, almost seven times the urban poverty rate. Such differences in poverty levels indicate very different poverty dynamics in the urban and rural areas of Vietnam. Further, the low poverty rates in the urban areas make analyses of

the poverty dynamics, based on the relatively small urban sample, highly uncertain because of the sampling variation. Therefore, we have chosen to limit the analysis to the rural parts of Vietnam (denoted Rural Vietnam).

The analysis is based on data from the panel component of the two household surveys VHLSS 2002 and VHLSS 2004, which includes 3,146 rural households. Each of the two household surveys is representative for Vietnam as a whole and for the urban and rural areas. However, it is uncertain if the rural household panel is representative for the rural area. In the following we assume it *is* representative and we estimate population parameters using the VHLSS 2002 sampling weights.

The classification of people into poor and non-poor is based on the World Bank/GSO definition of poverty. This means that per capita consumption expenditures are used to measure the individual's living standards and further that the poverty line is based on the Cost of Basic Needs approach.<sup>105</sup> The poverty line for 2002 is VND 1,916,672 per person per year (January 2002 prices) while it is VND 2,077,210 per person per year in 2004 (January 2004 prices). The increase in the poverty line reflects the inflation from 2002 to 2004.

Poverty transitions are defined in a straightforward way based on the observations of the households and individuals in 2002 and 2004. All individuals living in households in which the per capita consumption expenditure is below the poverty line in 2002 are classified as poor in 2002 while all other individuals are classified as non-poor, and likewise for 2004. The classification into poor and non-poor in the two years defines four transitions (or states): (i) being poor in both years (P–P), (ii) escaping poverty (P–NP), (iii) falling into poverty (NP–P) and (iv) being non-poor in both years (NP–NP). Hence, the four transitions disaggregate the change in the poverty rates from 2002 to 2004 into the gross movements into and out of poverty during this two year period.

Figure 1 is a cross-plot of per capita consumption in 2002 and 2004 (using log-scales). The two lines in the plot are the poverty lines for each of the years. These lines divide the sample of households into four sub-samples according to the classification given by the poverty transitions. The figure shows an average tendency of increasing per capita consumption expenditure (also considering the inflation in the period) and, more importantly, a fairly large concentration of observations around the poverty lines. In particular, many of the people moving out of poverty between 2002 and 2004 are just below the poverty line in 2002. This illustrates that, although average living standards has increased from 2002 to 2004, the observed poverty dynamics are highly dependent on the precise location of the poverty lines.

<sup>&</sup>lt;sup>105</sup> See World Bank (1999) for a thorough description of the computation of the per capita consumption expenditure measure and the derivation of the poverty line.



Figure 1: Real Per Capita Consumption Expenditure in 2002 and 2004

In Table 2 we report estimated poverty transitions for Rural Vietnam in the first row while the remainder of the table examines the transition patterns for households with different characteristics. The first column shows the share of the population that is endowed with the specific characteristic. The second and third columns report the shares of the population with the specific characteristic living in poverty in 2002 and 2004, respectively. Analogously, the last four columns show percentages of the relevant populations in each of the four poverty states specified above.

	Pop. Poverty			Poverty transitions 2002-2004				
	share	2002	2004	P - P	P - NP	NP- P	NP - NP	
Total	100.0	34.8	24.3	19.6	15.3	4.8	60.4	
Location								
Northern Uplands	15.5	21.3	24.7	26.1	15.1	19.0	11.9	
Red River Delta	20.1	15.0	12.0	11.1	20.1	15.5	23.4	
North Central Coast	16.4	22.9	23.4	25.4	19.6	15.1	12.7	
South Central Coast	8.5	7.3	7.7	6.2	8.6	14.0	8.8	
Central Highlands	6.8	12.4	12.4	14.1	10.1	5.4	3.8	
South East	9.1	3.9	2.7	1.5	6.9	7.6	12.2	
Mekong River Delta	23.6	17.3	17.0	15.5	19.6	23.5	27.2	
Gender of household head								
Female	15.1	12.5	13.1	11.7	13.6	19.0	16.2	
Male	84.9	87.5	86.9	88.3	86.4	81.0	83.8	
Education of household head								
None	33.6	42.9	47.8	50.1	33.6	38.4	27.8	
Primary school	25.8	26.7	25.9	26.4	27.2	23.8	25.4	
Lower secondary school	29.3	25.7	21.9	20.0	33.0	29.7	31.3	
Upper secondary school	6.8	3.5	3.9	3.5	3.6	5.8	8.7	
Technical/university	4.6	1.2	0.5	0.0	2.6	2.3	6.8	
Occupation of household head								
Agriculture	57.2	69.9	72.1	74.0	64.7	64.3	49.3	
Production	14.8	11.2	11.3	10.7	11.8	13.9	16.9	
Sales	7.3	4.5	3.6	3.6	5.7	3.8	9.2	
White-collar	9.0	4.6	3.5	2.7	7.1	6.8	11.7	
Not working	11.7	9.7	9.4	9.0	10.7	11.2	12.8	
Characteristics of dwelling								
Main source of lighting is not electricity	15.2	28.2	31.6	36.2	18.0	12.8	7.9	
Main source of lighting is electricity	84.8	71.8	68.4	63.8	82.0	87.2	92.1	
No access to clean water	63.0	79.9	79.5	82.6	76.4	66.7	52.9	
Access to clean water	37.0	20.1	20.5	17.4	23.6	33.3	47.1	
Commune characteristics								
No permanent road	55.2	65.2	65.6	68.1	61.5	55.1	49.4	
Permanent road	44.8	34.8	34.4	31.9	38.5	44.9	50.6	
No post office	56.9	63.6	63.9	64.4	62.6	61.6	52.6	
Post office	43.1	36.4	36.1	35.6	37.4	38.4	47.4	
No market	45.5	54.0	54.1	55.0	52.8	50.3	40.2	
Market	54.5	46.0	45.9	45.0	47.2	49.7	59.8	
No lower secondary school	65.6	68.5	66.7	66.3	71.2	68.2	63.8	
Lower secondary school	34.4	31.5	33.3	33.7	28.8	31.8	36.2	
No extension centre	71.7	74.3	73.6	74.7	73.8	69.3	70.4	
Extension centre	28.3	25.7	26.4	25.3	26.2	30.7	29.6	
Ethnicity of household head								
Kinh or Hoa	85.1	69.2	63.6	58.0	83.2	85.5	94.4	
Ethnic minority	14.9	30.8	36.4	42.0	16.8	14.5	5.6	

Table 2: Poverty and Poverty	Transition	Estimates	for Rural	Vietna	m

Notes: The four state indicators, P – P; P – NP; NP – P; NP – NP are; Poor in 2002 and 2004; Poor in 2004, Non-poor in 2004; Non-poor in 2002, poor in 2004 and Non-poor in 2002 and 2004.

Source: All figures are population estimates based on the panel data in VHLSS 2002 and 2004.

The panel-based estimates of the rural poverty rates are 34.9 percent in 2002 and 24.4 percent in 2004. These estimates correspond well with the official figures reported in Table 1. The 2002 poverty rate of 34.9 percent can be split into the fraction that remained poor in 2004 (19.6 percent) and the fraction that escaped poverty (15.3 percent). With 15.3 percent of the rural population escaping poverty from 2002 to 2004 it is clear that Vietnam is still, in the 2000s, extraordinary in terms of moving people out of poverty. Looking at the reverse movement, around 5 percent of the rural population fell into poverty by being non-poor in 2002 but poor in 2004. Although this is a fairly small fraction it underlines that pro-poor policies in Vietnam must now include the vulnerable households with per capita consumption expenditures just above the poverty line in order to keep the pace in poverty reduction.<sup>106</sup>

Location is an important factor both for the level of poverty and for the poverty transitions. The strong growth in Ho Chi Minh City, which is the urban centre in the South East, has a significant impact on the rural poverty in the region. While 9.1 percent of the rural population lives in the South East, the region has only 3.9 percent of the poor in 2002 and this share decreases further to 2.7 percent in 2004, corresponding to a regional poverty rate of 7.3 percent. However, the movement out of poverty is not as pronounced anymore as only 6.9 percent of the people escaping poverty in the period are from this region. This slower pace in poverty reduction is counteracted by a low share of the population falling into poverty. Hence, the region with the lowest poverty rate has relatively small gross movements in and out of poverty. The two deltas, the Red River Delta in the north and the Mekong River Delta in the south, are the most populous areas in Rural Vietnam. Both regions have poverty rates well below the average in both years and the shares of chronic poor (i.e., poor in both years) are also small relative to the population shares. In terms of people escaping poverty and falling into poverty, the Red River Delta has the best performance. The fourth region in Vietnam with below average poverty in 2002 and 2004 is the South Central Coast. Here the decline in poverty from 2002 to 2004 is modest and Table 2 shows that the reason for the modest performance is a relatively large group of people falling into poverty, while the share of people escaping poverty corresponds to the population share. The largest decline in the poverty rate is found in the Central Highlands where 22.5 percent of the population escaped poverty, corresponding to 10 percent of all who escaped poverty, whilst only 3.8 percent fell into poverty. However, this achievement must be seen against the background that the region has, by far, the highest poverty rate both in 2002 and in 2004. The Northern Uplands is also a critical region as the poverty rates are high, and more than one-quarter of the people being poor in both years live in this region. Further, jointly with the South Central Coast this region have high shares of people falling into poverty.

Turning to the other household characteristics, Table 2 shows that gender is not directly related to poverty rates and transitions in Vietnam. Individuals in households with a female household head have lover poverty rates compared to individuals in households headed by a male and there is no statistically significant difference between the poverty transitions even though the decline in the poverty rate from 2002 to 2004 is slightly lower for people in female headed households. However, many female headed households may be vulnerable in the sense that the per capita consumption expenditure is just above the poverty line. We find some indication of this in the table as relatively many female headed households fall into poverty in the period.

Education, measured as the highest level of formal education obtained by the household head in 2002, is strongly associated with poverty levels and dynamics as poverty rates are inversely related to the level of education both in 2002 and 2004. One interesting development is that poverty amongst

<sup>&</sup>lt;sup>106</sup> About half of the individuals who move from non-poor to poor have per capita consumption expenditures in 2002 that are less than 15 percent above the 2002 poverty line and around 90 percent of the individuals have per capita consumption expenditures less than 65 percent above the poverty line. Hence, although not poor, these individuals were clearly vulnerable in 2002.

people in households in which the head has completed a lower secondary school education has converged towards the poverty rate amongst the much smaller group of people in households headed by a person with higher secondary school education. To the extent that the association between education and poverty can be established as a causal relationship this points towards a possible propoor policy of motivating people to advance from primary to lower secondary schooling in the rural areas of Vietnam. The link from education to poverty may not be direct though, as education is correlated with the occupation of the household head and the latter is also associated with poverty.<sup>107</sup> In particular, individuals living in households in which the household head's occupation in 2002 is white-collar, sales or production have lower poverty rates in both years compared to individuals living in household head is working in agriculture. Moreover, there are smaller movements in and out of poverty for people having these non-farm jobs.

In addition to the education and occupation of the household head, characteristics of the dwelling and the local infrastructure are often found to be strong indicators of poverty. This is confirmed in Table 2 as we find that households in which the main source of lightning is not electricity have very high poverty rates and a high rate of people being poor in both years. However, we also find (a weak) sign of convergence in the poverty rates during the period both because relatively more people from households without electricity are moving out of poverty and because relatively less people from such households are falling into poverty. The importance of access to clean water is also clear, although the difference in poverty levels is not as extreme.<sup>108</sup>

Living in areas having a permanent road, a post office and/or a market is also indicative of a lower probability of being poor. But the table illustrates that the poverty rates are converging as the areas without these infrastructures in 2002 have larger decreases in the poverty rates than the areas in which the infrastructures are present.<sup>109</sup> Another interesting result is that being close to a lower secondary school does not appear important for the level of poverty. However, it is associated with the gross flows. Specifically, in areas without a lower secondary school the movements in and out of poverty appear to be larger than in areas with a lower secondary school.

Finally, looking at one of the most discussed correlates with poverty, ethnicity, we find large differences between the majority group consisting of ethnic Kinh and Hoa and the minority group (consisting of 52 different ethnicities). The panel data estimates show that the ethnic minority group constitutes just below 15 percent of the rural population, yet the share of the rural poor population is more than 30 percent, corresponding to an ethnic poverty rate of 72.5 percent in 2002, and among the chronic poor the share is above 40 percent, corresponding to 55 percent of the ethnic minority population. The poverty transitions show that the ethnic minority share of the populations moving out of and into poverty roughly corresponds to the population shares, hence, the convergence of majority and minority poverty rates is very slow.

The striking difference between the ethnic majority and minority groups underlines the need for detailed analyses of each group. Therefore we present poverty transition tables for the two groups in Tables 3 and 4. Table 3 covers the rural Kinh-Hoa population. The most salient difference between the whole rural population and the rural Kinh-Hoa population is that poverty is more equally distributed across location, education, occupation and infrastructural characteristics within the Kinh-Hoa population. Moreover, excluding the Northern Uplands, there are stronger signs of convergence in

<sup>&</sup>lt;sup>107</sup> In distinguishing between education and occupation one of the problems is that more than 80 percent of the household heads who have no formal education are either working in agriculture or not working. This share is declining steadily with the educational level of the household head although it is still above 60 percent for household heads with lower or upper secondary schooling.

<sup>&</sup>lt;sup>108</sup> Note, however, that the share of households with no access to clean water is more than twice the share of households with no electricity.

<sup>&</sup>lt;sup>109</sup> This convergence may be caused by investments in local infrastructure between 2002 and 2004.

poverty rates across regions as more people from the relatively poorer areas escape poverty. The Northern Uplands are special as the Kinh-Hoa poverty rate is below the national average in 2002 while the decrease in Kinh-Hoa poverty in that region is so small from 2002 to 2004 that the poverty rate is a little above average in 2004. The main reason for this development is a relatively large number of people falling into poverty in the period.

	Pop.	Pop. Poverty			Poverty transitions 200			
	share	2002	2004	P - P	P - NP	NP- P	NP - NP	
Total	100.0	28.3	18.1	13.3	14.9	4.8	67.0	
Location								
Northern Uplands	9.2	8.3	10.1	9.6	7.0	11.6	9.4	
Red River Delta	23.4	21.1	18.0	17.9	23.9	18.2	24.7	
North Central Coast	17.6	28.7	31.0	35.9	22.2	17.4	12.9	
South Central Coast	9.4	8.8	9.7	7.3	10.2	16.2	9.2	
Central Highlands	4.7	7.5	5.9	6.9	8.1	3.2	3.7	
South East	9.8	4.0	2.8	1.0	6.7	7.9	12.4	
Mekong River Delta	25.8	21.6	22.5	21.3	21.9	25.6	27.6	
Gender of household head								
Female	16.4	15.7	16.6	15.2	16.1	20.4	16.4	
Male	83.6	84.3	83.4	84.8	83.9	79.6	83.6	
Education of household head								
None	30.3	35.9	39.8	40.3	31.9	38.2	27.3	
Primary school	25.9	27.1	26.2	27.7	26.7	21.9	25.6	
Lower secondary school	31.6	31.9	28.0	27.1	36.2	30.4	31.5	
Upper secondary school	7.4	4.3	5.3	4.7	3.9	6.8	8.8	
Technical/university	4.9	0.8	0.8	0.1	1.4	2.7	6.8	
Occupation of household head								
Agriculture	52.6	62.1	61.9	62.1	62.1	61.2	47.9	
Production	17.0	15.8	17.2	17.6	14.2	16.1	17.5	
Sales	8.1	5.1	4.5	4.6	5.5	4.3	9.6	
White-collar	9.3	4.3	3.5	2.3	6.1	7.0	11.5	
Not working	13.1	12.7	12.8	13.3	12.2	11.4	13.3	
Characteristics of dwelling								
Main source of lighting is not electricity	9.7	15.6	17.5	19.9	11.8	10.7	7.1	
Main source of lighting is electricity	90.3	84.4	82.5	80.1	88.2	89.3	92.9	
No access to clean water	58.9	75.5	74.2	78.0	73.3	63.4	51.5	
Access to clean water	41.1	24.5	25.8	22.0	26.7	36.6	48.5	
Commune characteristics								
No permanent road	52.4	60.4	59.3	62.1	58.9	51.6	49.0	
Permanent road	47.6	39.6	40.7	37.9	41.1	48.4	51.0	
No post office	53.5	57.4	56.4	55.6	59.0	58.5	51.5	
Post office	46.5	42.6	43.6	44.4	41.0	41.5	48.5	
No market	41.6	47.0	46.6	46.0	47.9	48.3	38.9	
Market	58.4	53.0	53.4	54.0	52.1	51.7	61.1	
No lower secondary school	65.0	68.9	65.7	65.1	72.2	67.3	63.2	
Lower secondary school	35.0	31.1	34.3	34.9	27.8	32.7	36.8	
No extension centre	70.4	70.6	69.4	68.9	72.1	70.9	70.3	
Extension centre	29.6	294	30.6	31.1	27.9	29.1	297	

|--|

*Notes:* The four state indicators, P – P; P – NP; NP – P; NP – NP are; Poor in 2002 and 2004; Poor in 2004, Non-poor in 2004; Non-poor in 2002, poor in 2004 and Non-poor in 2002 and 2004.

Source: All figures are population estimates based on the panel data in VHLSS 2002 and 2004.

	Pop.	Po۱	/erty	Pover	ty transiti	ons 200	2-2004
	share	2002	2004	P - P	P - NP	NP- P	NP - NP
Total	100.0	72.5	59.9	55.2	17.3	4.6	22.9
Location							
Northern Uplands	51.5	50.3	49.9	48.9	54.8	62.7	53.1
Red River Delta	1.6	1.5	1.5	1.6	1.3	0.0	1.9
North Central Coast	9.5	10.0	10.2	10.9	7.0	1.6	9.4
South Central Coast	3.2	3.7	4.4	4.7	0.5	1.1	2.0
Central Highlands	18.8	23.1	23.7	24.1	19.7	18.4	5.3
South East	4.7	3.7	2.6	2.4	8.1	5.6	7.8
Mekong River Delta	10.7	7.7	7.6	7.4	8.7	10.5	20.4
Gender of household head							
Female	7.5	5.6	7.2	6.9	1.2	10.7	13.2
Male	92.5	94.4	92.8	93.1	98.8	89.3	86.8
Education of household head							
None	52.4	58.5	61.8	63.7	41.8	39.7	35.9
Primary school	25.3	25.8	25.4	24.6	29.9	34.9	21.6
Lower secondary school	16.3	11.8	11.2	10.0	17.5	25.4	28.6
Upper secondary school	2.9	1.8	1.6	1.8	1.8	0.0	7.2
Technical/university	3.1	2.1	0.0	0.0	9.0	0.0	6.6
Occupation of household head							
Agriculture	83.5	87.3	89.7	90.3	77.7	82.5	71.5
Production	2.1	0.8	1.1	1.1	0.0	1.1	6.2
Sales	3.0	3.3	2.1	2.2	6.8	1.1	2.7
White-collar	7.6	5.5	3.5	3.3	12.2	5.4	15.0
Not working	3.8	3.1	3.6	3.1	3.3	10.0	4.7
Characteristics of dwelling							
Main source of lighting is not electricity	47.0	56.2	56.0	58.6	48.6	25.0	22.1
Main source of lighting is electricity	53.0	43.8	44.0	41.4	51.4	75.0	77.9
No access to clean water	86.3	89.5	88.7	88.9	91.5	86.1	75.9
Access to clean water	13.7	10.5	11.3	11.1	8.5	13.9	24.1
Commune characteristics							
No permanent road	71.5	76.1	76.4	76.5	74.7	75.8	56.1
Permanent road	28.5	23.9	23.6	23.5	25.3	24.2	43.9
No post office	76.2	77.5	76.9	76.6	80.1	79.9	71.5
Post office	23.8	22.5	23.1	23.4	19.9	20.1	28.5
No market	67.7	69.7	67.1	67.5	76.8	62.4	62.7
Market	32.3	30.3	32.9	32.5	23.2	37.6	37.3
No lower secondary school	69.2	67.6	68.5	68.1	66.1	73.5	73.2
Lower secondary school	30.8	32.4	31.5	31.9	33.9	26.5	26.8
No extension centre	78.8	82.5	80.8	82.6	82.3	59.6	70.8
Extension centre	21.2	17.5	19.2	17.4	17.7	40.4	29.2

#### Table 4: Poverty and Poverty Transition Estimates for the Ethnic Minorities in Rural Vietnam

 Extension centre
 21.2
 17.5
 19.2
 17.4
 17.7
 40.4
 29.2

 Notes: The four state indicators, P - P; P - NP; NP - P; NP - NP are; Poor in 2002 and 2004; Poor in 2004, Non-poor in 2004; Non-poor in 2002, poor in 2004 and Non-poor in 2002 and 2004.
 2002
 2004; Poor in 2004, Non-poor in 2004; Non-poor in 2004; Non-poor in 2004; Non-poor in 2004, Non-

Source: All figures are population estimates based on the panel data in VHLSS 2002 and 2004.

Within the rural ethnic minority population differences in the household characteristics appear to be more important than to the whole rural population and, hence, also compared to the rural Kinh-Hoa population—in particular when looking at the movements in and out of poverty (Table 4). Even though ethnic poverty in 2002 and 2004 is in general more evenly distributed across the geographical regions than the overall rural poverty, the movements in and out of poverty are strongly associated with the regions. More than half of the rural minority population lives in the Northern Uplands where the minority poverty rate is below the national average both in 2002 and 2004. In this region the flows in and out of poverty are above the average, much in line with the findings for the Kinh-Hoa majority. The movements into and out of poverty is much larger than the inflow resulting in a fairly large decline in the poverty rate in that region. In contrast, the movements in and out of poverty rate, resulting in a fairly large to the south Central Coast, which is the region with the highest rural minority poverty rate, resulting in a very small decline in the poverty rate from 2002 to 2004 and in the fact that most of the poor minorities in this region are chronic poor.<sup>110</sup>

Individuals living in female headed households appear to have different poverty transitions relative to individuals living in male headed households and this point towards a possible gender inequality problem the minority population in the future. However, the share of female headed households in the minority population is so small that we cannot make any conclusions based on the panel sample of the minorities. In contrast, the differences in the transitions in and out of poverty classified by the occupation of the household head are large enough to conclude that individuals living in households in which the head is employed in sales and white-collar occupations have done much better than the other groups in terms of escaping poverty and not falling into poverty. The same goes for the classification by education where we find large effects of secondary schooling and the remarkable result that none of the (few) minority households with technical and university education are poor in 2004.

Finally, we do not find statistically significant differences in the poverty transitions for the minorities when looking at most of the indicators for local infrastructure. Apart from the presence of a road, none of the indicators have a significant association with the poverty transitions. This surprising result may be due to the relatively small size of the rural minority sample.

Moving beyond the differences in poverty dynamics governed by various categorical classifications we focus, in Table 5, on the agricultural production and household assets across poverty states. As shown in Table 2, more than half of the rural household heads are working in agriculture and even if the household head is not working in agriculture the household is often engaged in agricultural production. Therefore it is of interest to look at the distribution of the value of the agricultural production.

<sup>&</sup>lt;sup>110</sup> The incidence of chronic poor rural minorities in the South Central Coast is 81 percent.

		P - P	I	P - NP		NP - P	Ν	P - NP
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
	Rural Vi	etnam						
Rice	5.09	4.41	6.51	5.10	6.94	5.16	10.37	5.96
Coffee	0.49	0	0.96	0	0.54	0	0.82	0
Fruit	0.41	0.08	0.80	0.08	1.81	0.06	2.77	0
Pigs	1.95	1.21	2.89	1.80	3.10	2.48	5.72	2.50
Feathered animals	0.67	0.40	0.85	0.52	0.96	0.46	1.57	0.41
Aquaculture	1.04	0	1.83	0	1.79	0	4.58	0
Productive assets	8.75	4.40	12.29	6.29	18.87	6.99	39.22	18.06
	Kinh-Ho	а						
Rice	5.40	4.68	6.62	5.22	6.73	4.67	10.43	5.90
Coffee	0.46	0	1.11	0	0.59	0	0.80	0
Fruit	0.38	0	0.82	0.02	1.94	0.02	2.80	0
Pigs	2.32	1.74	2.99	2.15	2.88	1.78	5.74	2.42
Feathered animals	0.57	0.37	0.87	0.52	0.89	0.40	1.56	0.37
Aquaculture	1.37	0	2.09	0	1.80	0	4.76	0
Productive assets	8.87	3.69	11.89	6.09	19.75	7.66	39.36	17.89
	Ethnic N	linorities						
Rice	4.66	4.13	6.01	4.68	8.20	6.02	9.42	6.34
Coffee	0.52	0	0.22	0	0.28	0	1.09	0
Fruit	0.47	0.16	0.67	0.14	1.02	0.30	2.34	0.24
Pigs	1.44	0.91	2.42	1.36	4.39	2.82	5.49	2.94
Feathered animals	0.80	0.48	0.75	0.59	1.34	0.84	1.86	1.02
Aquaculture	0.59	0.14	0.55	0.19	1.69	0	1.49	0.18
Productive assets	8.59	5.04	14.28	7.05	13.69	6.60	36.77	21.76

Table 5: Means and Medians of the Per Capita Value of Agricultural Production and
Productive Assets in 2002, by Poverty States (100,000 VND)

*Notes*: The four state indicators, P – P; P – NP; NP – P; NP – NP are; Poor in 2002 and 2004; Poor in 2004, Non-poor in 2004; Non-poor in 2002, poor in 2004 and Non-poor in 2002 and 2004.

Source: All figures are population estimates based on the panel data in VHLSS 2002 and 2004.

In Table 5 we show the estimated means and medians of the per capita monetary value of the production in 2001/2002 of different agricultural commodities and of the per capita monetary value of productive assets in 2002 across the four poverty states.<sup>111,112</sup> Starting with the value of rice production, the main agricultural commodity in rural Vietnam, we find an increasing mean (and median) as we move from the chronic poor to the 'never poor'. The mean per capita value of rice production in the group of never poor individuals is twice as large as the mean value in the group of chronic poor. Yet, comparing the means for the individuals who escape poverty and the individuals who fall into poverty we find a small and statistically insignificant difference. Hence, the value of rice production is dividing the population into three segments; the chronic (or very) poor, the vulnerable who are close to the poverty line, and the more affluent who are never poor. This segmentation appears to be typical for the production value of agricultural commodities as we find statistically insignificant differences between the group of individuals falling into poverty while we generally find significant differences between the chronic poor and the individuals escaping poverty and between the

<sup>&</sup>lt;sup>111</sup> The per capita monetary values (VND 100,000; January 2002) reflect the production values a household generated during the last twelve months prior to the 2002 interview divided by the household size.

<sup>&</sup>lt;sup>112</sup> Productive assets are defined as household asset items, which the household can use in production activities. Hence, these assets do not include durable goods such as televisions, radios and air-conditioners.

never poor and the individuals falling into poverty. The main exception is coffee production for which we cannot reject the hypothesis of equal per capita mean values across the four groups. For fruit and pig production and aquaculture we find more skewed distributions than for rice production in the sense that the ratio of the means of the chronic poor and never poor is larger than two. However, in terms of absolute values rice production is clearly dominating the picture.

Breaking down the per capita values of agricultural production by ethnicity does not change the overall picture much. The estimated means for the Kinh-Hoa subpopulation are close to the full sample estimates and we find the same pattern of segmentation into three groups; the very poor, the vulnerable and the more affluent. For the ethnic minorities we do not find significant differences in the means between neighboring groups, probably because of the small sample sizes, but we do find significant differences between the chronic poor and the never poor—except for the value of coffee production. One interesting difference between the Kinh-Hoa and the ethnic minorities is that the value of rice production appears to be smaller, on average, in the chronic poor minority group compared to the chronic poor Kinh-Hoa. In contrast the ethnic minorities who fall into poverty seem to have a higher value of rice production than the group of Kinh-Hoa who fall into poverty.

Finally, looking at the value of productive assets in 2002, we find a statistically significant ranking of the four poverty groups in the full sample. But in the ethnic minority sample the estimated mean in the group who fall into poverty is actually smaller than the estimated mean in the group who escape poverty. The difference is, however, not statistically significant.

Overall, the results reported in Tables 2-5 indicate that the poverty dynamics from 2002 to 2004 can be associated with a large set of systematic determinants such as location, education, occupation, agricultural production and household assets. It is, however, not clear if all of the determinants are directly affecting the poverty dynamics or if some of the observed associations are working through other channels. We have already discussed the difficulty in establishing a direct causal effect of education because of the association between education and non-farm occupation. But there is also a strong link between location and occupation as non-farm jobs are not equally distributed across the seven regions. Hence, there is a need for a more systematic analysis of the variation in the poverty dynamics. We report results of such an analysis in the next section.

#### 3. An Econometric Analysis of the Poverty Transitions

In the joint analysis of the many factors that may have an impact on the poverty transitions from 2002 to 2004 we follow the previous studies of poverty transitions in Vietnam (e.g., Glewwe, Gragnolati and Zaman, 2002; Justino and Litchfield 2003; Litchfield and Justino, 2004; Nimi, Vasudeva-Dutta and Winters, 2003, 2004) and formulate a multinomial logit model for the four poverty states defined in Section 2. Hence, we assign the numbers 1-4 to the four poverty states 'P-P', 'P-NP', 'NP-P', and 'NP-NP' and let *y* denote a random variable taking on one of the four values. The multinomial logit model is then used to estimate the response probabilities for an individual *i*, conditional on a set of variables,  $x_i$  (given as a 1 x *K* vector). The model is specified as<sup>113</sup>

$$p_{ij} = \Pr(y_i = j \mid x_i) = \frac{\exp(x_i \beta_j)}{\sum_{l=1}^{4} \exp(x_l \beta_l)}, \quad j = 1, 2, 3, 4.$$
(1)

The model is estimated using the pseudo-likelihood procedure that takes account of the stratification and clustering of the household surveys and we use the VHLSS 2002 sampling weights to get

<sup>&</sup>lt;sup>113</sup> The probabilities sum to one, so we need to impose an identifying restriction on the parameters. The most common restriction is to set  $\beta_1 = 0$ . However, this is not important for the presentation and interpretation of the empirical results.

population estimates of the parameters.<sup>114</sup> In the regressions we omit all individuals living in households in which the head and/or the spouse has a technical/university degree because the frequency of such households being poor in both years is very close to zero, and because the poverty rate for the ethnic minority households with this educational endowment is zero in 2004 implying that none of these households are poor in both periods or fall into poverty from 2002 to 2004. This restriction leaves 2957 observations (1161 primary sampling units) in the estimation sample. Further, we include information about household demographics such as the size of the household (in logs), the share of children (below 19 years of age), the share of elderly household members (female adults above 55 and male adults above 60 years) and the age of the household head. All explanatory variables in the model are from the 2002 household survey. Hence, we do not include any information after 2002 apart from the poverty status in 2004. The reason for omitting explanatory variables from the 2004 survey is that we wish to keep away from predicting the 2002 poverty status using future information in order to avoid reverse causality problems.

Figure 2 shows the frequency distributions of the estimated conditional probabilities for each of the four poverty transition states. As seen, the model lacks precision in correctly classifying the two cross-over states in the sense that the estimated probabilities of escaping poverty (P-NP) and, in particular, of falling into poverty (NP-P) are quite low for all individuals in the sample. Still, based on standard measures of fit, such as McFadden's pseudo  $R^2$  (= 0.28), the model has reasonable explanatory power. Further, the null-hypotheses of equality of all slope parameters between any pair of states are clearly rejected.





<sup>&</sup>lt;sup>114</sup> The model is estimated using STATA versions 9 and 10.

We present and discuss the parameter estimates in two different ways. In Table 6 we report the percentage change in relative risks which is derived from the relative risk ratio. The relative risk ratio between states j and l is the risk of being in state j rather than state l and for the multinomial logit model it is defined as

$$\operatorname{RRR}_{jl} = \frac{p_{ij}}{p_{il}} = \frac{\exp(x_i\beta_j)}{\exp(x_i\beta_l)} = e^{x_i(\beta_j - \beta_l)}, \quad j, l = 1, \dots, 4.$$
<sup>(2)</sup>

Based on this measure we compute the percentage change in the relative risk as

$$DRR_{jlk} = 100 \times (e^{x_{ik}(p_{jk} - p_{lk})} - 1), \quad j, l = 1, \dots, 4, k = 1, \dots, K.$$
(3)

When we consider a one unit change in  $x_{ik}$ , the relative risk ratio reduces to  $\exp(\beta_{ik}-\beta_k)$  and, as seen, the advantage of looking at the relative risk ratio and the percentage change in the relative risk is that these measures are independent of the levels of the other conditioning variables. In Table 6 we focus on the change in the relative risk of escaping poverty versus being poor in both years (Pr(P-NP) / Pr(P-P)), the risk of falling into poverty versus being non-poor in both periods (Pr(P-NP) / Pr(NP-NP)) and, for completeness, the chance of being non-poor in both periods relative to being poor in both periods (Pr(NP-NP) / Pr(P-P)).

In addition to the changes in the relative risks we also look at the marginal effects on the conditional probabilities of a change in the regressors (Table 7). The marginal effects are estimated from

$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij}(\beta_j - \sum_{l=1}^4 p_{ll}\beta_l).$$
(4)

As these effects depend on the probabilities of being in each of the poverty states and, thus, on the values of the conditioning variables we aim at evaluating the effects for a base individual who is not too far from the poverty lines and, further, is not too 'rare' in the sample. Hence, our base individual lives in a household in the Mekong River Delta. The household, which consist of 3 adults—none of which are elderly-and 2 children, is headed by a 44 year old Kinh-Hoa male with no formal education who is working in agriculture. The dwelling, which has electricity but no access to clean water, is located in a commune that has neither a permanent road, a post office, a market, a lower secondary school, nor an agricultural extension center. The values of the agricultural commodities and the value of productive assets are determined by looking at the median values in households with 2002 per capita consumption expenditures within a band 25 percent above and below the poverty line in 2002. The specific value for each of the regressors is reported in the last column of Table 7 and the estimated conditional probabilities are reported in the bottom row of the table. Compared to the unconditional probability estimates (the transition shares reported in Table 2) the base individual has much higher probabilities of escaping poverty and falling into poverty (almost a doubling in both cases) and these higher probabilities imply a lower probability of being non-poor in both periods whereas the probability of being poor in both periods is close to the unconditional estimate. We consider this to be a relevant base for analyzing the marginal effects.

Turning to the empirical results, one of the most interesting findings is the insignificance of the commune characteristics. Apart from the presence of a permanent road, none of the included commune characteristics are important in determining the poverty transitions of the rural population. This result is in contrast with the analyses of poverty transitions in the 1990s based on the 1992 and 1998 VLSS in Glewwe *et al.* (2002) where similar regressions indicate that the presence of a post office and/or a market had a significant impact on the poverty transitions.<sup>115</sup> In addition, we find that even

<sup>&</sup>lt;sup>115</sup> However, the result for the presence of a post office is surprising in Glewwe et al. as this indicates a higher probability of being poor in both 1992 and 1998 rather than escaping poverty in the period. The authors note this but they do not directly explain the result. See Glewwe et al. (2002, pp. 789-90).

though the presence of a permanent road is a significant determinant of the poverty transitions as such, it has no significant impact on the chance of escaping poverty relative to being poor in both periods or the risk of falling into poverty relative to being non-poor in both periods. Hence, the presence of a permanent road in 2002 mainly separates the poor from the non-poor in 2002 while it carries little or no information about movements in and out of poverty. However, our results should not be interpreted directly as a finding of lack of importance of local infrastructure for poverty reduction. As our model does not distinguish between older and newly established infrastructural units and, further, does not include units established between 2002 and 2004, we cannot infer that improving the infrastructure in a commune has no impact on poverty. A more careful analysis is needed in order to make robust conclusions about the importance of local infrastructure in the rural areas of Vietnam.<sup>116</sup>

		Pr(P	– NP <u>)</u>			<u>Pr(NP</u>	– P <u>)</u>		P	r(NP	– NP)	
		Pr(P	– P)			Pr(NP -	- NP)			Pr(P	– P)	
	DRR		min95	max95	DRR		min95	5 max95	DRR		min95	max95
Ethnic minority	-60.7	***	-77.4	-31.9	59.1		-21.5	222.7	-75.2	***	-86.7	-53.7
Northern Uplands	-58.1	**	-79.0	-16.2	63.8		-26.4	264.4	-68.2	***	-83.9	-37.1
Red River Delta	-50.5	**	-74.1	-5.6	27.6		-42.5	183.1	-85.2	***	-91.9	-73.2
North Central Coast	-73.5	***	-85.9	-50.4	32.5		-39.2	188.8	-88.6	***	-93.8	-79.1
South Central Coast	-12.1		-56.5	77.6	46.8		-30.3	209.1	-32.4		-65.6	32.8
Central Highlands	-42.5		-73.4	24.3	19.2		-59.4	250.0	-74.2	***	-87.9	-44.7
South East	238.0	**	25.4	811.1	-43.1		-76.1	35.4	333.4	**	58.4	1086
Household characteristics												
Female head	2.3		-48.4	102.7	-6.3		-55.6	97.7	17.5		-38.3	124.1
Age of head	-0.4		-2.6	1.7	-4.0	***	-6.4	-1.6	4.5	***	2.3	6.7
Age of head, squared	0.0		-0.1	0.1	0.0		0.0	0.1	-0.1	**	-0.1	0.0
No spouse	25.6		-41.2	168.4	86.3		-16.2	314.2	-19.1		-60.4	65.2
Share of children <sup>1</sup>	-17.6		-36.2	6.5	57.9	***	15.3	116.1	-41.2	***	-53.8	-25.0
Share of elderly <sup>1</sup>	-3.6		-26.5	26.5	62.0	***	22.7	113.9	-39.4	***	-52.6	-22.4
Log of household size <sup>2</sup>	-0.9		-11.6	11.2	6.0	**	-5.4	18.8	-24.9	***	-32.6	-16.4
Head's education and occup	ation											
Primary school	42.6		-11.3	129.0	-39.4	**	-65.5	6.6	92.3	***	26.8	191.8
Low. secondary school	131.6	***	42.2	277.2	-38.7	*	-65.2	8.0	226.7	***	107.7	414.1
High. secondary school	51.4		-35.3	254.6	-47.4		-79.1	32.1	266.2	***	79.8	645.9
Production	-10.7		-47.3	51.6	-52.3	**	-74.5	-10.7	115.0	***	32.7	248.3
Sales	82.5		-16.5	299.2	-74.9	**	-91.7	-23.9	400.1	***	135.6	961.3
White-collar	142.1	**	8.7	439.0	-44.1		-76.2	31.0	335.7	***	86.6	917.3
Not working	4.9		-51.3	126.1	-35.9		-68.3	30.0	77.1	*	-6.6	235.5
Agricultural production												
Rice	3.1	**	0.6	5.6	-2.6	**	-4.8	-0.3	6.2	***	3.7	8.7
Coffee	1.6		-0.8	4.0	-1.5		-7.4	4.8	1.8		-1.8	5.5
Fruit	13.6	**	1.7	26.8	-0.5		-3.1	2.2	17.8	***	6.0	30.8
Pigs	6.5	**	0.7	12.7	-4.9	**	-8.9	-0.6	13.4	***	7.7	19.4
Feathered animals	18.1	*	-1.7	42.0	-6.1		-13.7	2.1	29	***	10.0	51
Aquaculture	3.2		-0.9	7.3	-1.3		-3.2	0.7	4.3	**	0.5	8.2

Table 6: Results of Multinomial Logistic Regression for Rural Vietn	am
(Percentage Change in Relative Risk)	

<sup>116</sup> See e.g., Mu and van de Walle (2007) for a comprehensive analysis of the impact of rural road rehabilitation in Vietnam.

	<u>Pr(P – NP)</u> Pr(P – P)					<u>Pr(NP – P)</u> Pr(NP – NP)					<u>Pr(NP – NP)</u> Pr(P – P)			
	DRR	,	, min95	5 max95	DRR	,	min95	5 max95	5 DRR	`	, min95	i max95		
Assets														
Productive assets	1.4	*	0.0	2.9	-0.8	*	-1.6	0.1	4.5	***	3.0	6.0		
No electricity	-40.3	**	-61.2	-8.3	21.6		-35.2	128.1	-72.7	***	-82.9	-56.5		
Access to clean water	-1.2		-37.5	56.2	-21.1		-50.1	24.7	131.6	***	51.3	254.5		
Commune characteristics														
Permanent road	30.2		-11.2	90.9	-8.1		-40.7	42.2	85.5	***	28.9	166.9		
Post office	-1.1		-35.4	51.5	-6.6		-41.5	49.1	0.7		-33.4	52.2		
Market	-18.8		-46.0	22.0	-25.0		-54.9	24.5	2.1		-32.3	54.0		
Lower secondary school	-15.2		-43.7	27.8	-7.6		-44.0	52.3	13.4		-25.7	72.9		
Agricultural extension center	-15.9		-42.3	22.5	8.0		-34.4	77.7	-4.6		-34.3	38.6		

*Notes:* min95 and max95 are the lower and upper 95% confidence bounds. \*, \*\* and \*\*\* indicates significance at the 10%, 5% and 1% level, respectively. <sup>1</sup>The effect is evaluated for a change in the share of children/elderly 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The effect is evaluated at a change in the household size from five to six. *Source:* Own calculations based on the VHLSS 2002 and 2004.

A finding that we do consider to be robust is the importance of ethnicity. Belonging to the ethnic minority instead of the Kinh-Hoa majority decreases the chance of escaping poverty by 61 percent, relative to being poor in both periods, even when we condition on location, household characteristics, education, occupation, the value of agricultural production etc. Moreover, the conditional risk of falling into poverty is almost 60 percent higher for the ethnic minorities relative to the Kinh-Hoa majority. In terms of changes in the conditional probabilities we find that if the baseline individual is from an ethnic minority household instead of a Kinh-Hoa household the probability of being poor in both periods increases from about 21 to 45 percent while the probabilities of being in any of the three other states decrease. The largest decrease is in the probability of being non-poor in both periods, which drops from 43 to 23 percent. The problem of increasing the welfare for the ethnic minority people in Vietnam is well known and discussed in depth elsewhere (see e.g., Chapter 5) and we look more into the poverty dynamics for the ethnic minority group below.

The results for location are in line with the unconditional results reported in Table 2. Thus, in the southern parts of Vietnam the chances of escaping poverty are significantly higher than in the northern parts, and this goes in particular for the area around Ho Chi Minh City; the South East Region. The estimates show that if the baseline individual lives in one of the regions north of the South Central Coast instead of in the Mekong River Delta the probability of being poor in both periods more than doubles. In contrast, location does not significantly affect the conditional risk of falling into poverty relative to being non-poor in both periods. This result is probably driven by the model's lack of precision in classifying the individuals falling into poverty.

The results are in some sense 'reversed' when we look at household demographics. Household size and composition does not significantly impact on the chance of escaping poverty relative to being poor in both periods, but household size and its composition in terms of the share of children and elderly household members carries information about the risk of falling into poverty. Each additional child or elderly, keeping the household size fixed, increases the risk of falling into poverty by around 60 percent. Moreover, households headed by a young person have a slightly higher risk of falling into poverty. This result shows that young families with many children, and three generation households, not only have a higher probability of being poor than older families with less kids they are also more vulnerable than the older families even when both families are initially above the poverty line.

	P - P	P - P		P - NP			NP - P			NP - N			
	∂piıl∂	<sup>7</sup> x	s.e.	∂pi₂lĉ	X	s.e.	∂різ/∂х		s.e.	∂pi4l ĉ	<sup>7</sup> X	s.e.	x
Ethnic minority	24.5	***	6.6	-4.0		4.6	-1.1		2.2	-19.5	***	5.2	0
Northern Uplands	20.3	***	6.7	-4.8		5.7	0.3		2.9	-15.7	***	6.1	0
Red River Delta	27.7	***	6.5	4.7		6.5	-4.5	*	2.4	-27.9	***	4.9	0
North Central Coast	39.5	***	5.8	-6.5		5.6	-4.5	*	2.4	-28.5	***	5.2	0
South Central Coast	4.3		5.7	1.9		5.9	1.6		3.2	-7.8		5.9	0
Central Highlands	20.1	***	7.4	3.9		7.3	-3.1		2.9	-20.9	***	6.3	0
South East	-14.2	***	4.5	1.4		6.4	-1.9		2.7	14.7	**	7.3	0
Household characteristics													
Female head	-1.7		4.8	-1.8		4.6	0.1		2.7	3.4		5.4	0
Age of head	-0.4	***	0.2	-0.6	***	0.2	-0.1		0.1	1.1	***	0.2	44
No spouse	-0.6		5.5	6.2		6.2	3.7		4.2	-9.3	*	5.5	0
Share of children <sup>1</sup>	5.9	***	2.1	2.7		2.2	1.7		1.2	-10.4	***	2.2	0.4
Share of elderly <sup>1</sup>	4.6	**	2.0	5.4	**	2.4	1.7		1.1	-11.7	***	2.4	0
Log of household size <sup>2</sup>	2.9	***	1.0	3.9	***	1.1	-0.7		0.5	-6.1	***	1.1	ln(5)
Head's education and occupation													
Primary school	-7.1	**	2.9	-2.0		4.0	-1.9		1.8	11.0	***	4.1	0
Low. secondary school	-12.1	***	3.2	-1.3		4.4	-1.4		1.9	14.8	***	4.6	0
High. secondary school	-11.8	***	3.7	-10.3	**	5.7	-1.5		2.9	23.6	***	6.3	0
Production	-6.5	**	3.3	-11.2	***	3.7	-2.4 '	*	1.9	20.1	***	4.3	0
Sales	-13.6	***	3.7	-11.1	**	4.9	-4.6		2.4	29.3	***	5.4	0
White-collar	-13.6	***	3.9	-5.2		5.8	-1.4		2.6	20.2	***	7.1	0
Not working	-5.4		4.2	-6.5		5.3	-1.3		2.3	13.1	**	5.6	0
Agricultural production													
Rice	-0.8	***	0.2	-0.2		0.2	0.0		0.1	1.0	***	0.2	6.2
Coffee	-0.3		0.2	0.1		0.2	-0.1		0.2	0.2		0.4	0
Fruit	-2.4	**	1.0	0.2		0.6	0.3	k	0.2	1.9	***	0.7	0.1
Pigs	-1.6	***	0.5	-0.4		0.4	0.0		0.2	2.0	***	0.4	2.5
Feathered animals	-3.5	***	1.3	-0.1		1.9	0.1		0.4	3.5	**	1.8	0.5
Aquaculture	-0.6	*	0.3	0.1		0.2	0.0		0.0	0.5	**	0.2	0
Assets													
Productive assets	-5.3	***	1.4	-3.3	**	1.3	0.8	*	0.5	7.8	***	1.3	8
No electricity	19.0	***	4.3	4.3		4.4	-2.9		1.9	-20.3	***	4.4	0
Access to clean water	-7.9	***	2.9	-11.3	***	3.1	1.0		1.9	18.2	***	3.7	0
Commune characteristics													
Permanent road	-6.9	***	2.4	-3.9		3.1	1.0		1.7	9.8	***	3.5	0
Post office	0.1		3.1	-0.2		3.5	-0.4		1.6	0.5		3.8	0
Market	1.4		3.1	-3.8		3.2	-1.5		1.6	3.9		4.0	0
Lower secondary school	-0.3		3.1	-4.8		3.1	0.2		1.8	4.9		4.0	0
Agricultural extension center	1.4		2.8	-2.9		3.1	0.8		1.9	0.8		3.7	0
<i>D</i> ij	20.5			28.7			8.0			42.8			

*Notes:* Probabilities are given as percentages and marginal effects as percentage points. \*, \*\*, and \*\*\* indicates significance at the 10%, 5% and 1% level of significance, respectively. Marginal effects of categorical variables are computed as the change in the estimated probability when the indicator variable changes from 0 to 1. The base values of the regressors are given in the last column. <sup>1</sup>The marginal effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The marginal effect is evaluated for a change in the household size from five to six.

Source: Own calculations based on the VHLSS 2002 and 2004.

Looking at the household head's education and occupation it is clear that the main impact of having completed an education or working outside agriculture is in the chance of being non-poor in both periods relative to being poor. This is reflected in many insignificant differences in the chances of escaping poverty relative to being poor in both periods and in the only marginally significant differences in the risk of falling into poverty. However, a few results do stand out. First of all we find the chance of escaping poverty is more than 100 percent higher for households in which the head has completed lower secondary school compared to households in which the head has no education. And as the chance of escaping poverty is more or less the same for households in which the head has either no education or completed primary school we can infer that moving from primary school to lower secondary school education increases the chance of escaping poverty. In addition, the risk of falling into poverty is lower when the household head has completed either primary school or lower secondary school (although these results are only marginally significant). As noted in Section 2, these results point towards a poverty reduction strategy that involves increasing the number of people in the rural areas who complete lower secondary school and this lends support to the Government of Vietnam's educational policy, which includes a plan of lowering the private cost of attending secondary school to increase the educational attainment (see SRV, 2003).

With respect to occupation, white-collar workers have a much higher chance of escaping poverty than agricultural workers (about 140%, with a large margin of error though). Although this result is not surprising as such, it is not clear from the unconditional results in Table 2 and it is different from the results for sales and production for which we do not find significant differences in the chance of escaping poverty relative to being poor in both periods. Instead, we find that when the household head is working in sales or production, the household members are less vulnerable in the sense that the risk of falling into poverty is lower than for members of households in which the head is working in agriculture. Overall, this leads us to conclude that the chances of becoming, or being, non-poor are significantly higher when the household head is working outside agriculture, supporting the idea of improving the options for non-farm jobs in the rural areas.

However, a non-farm job is not the only path out of poverty in rural Vietnam. Increased agricultural production, or higher prices, also increases the chance of being or becoming non-poor. In particular, higher values of rice, fruit, pig and (to some extent) feathered animal production in 2002 indicates better chances of escaping poverty. Further, rice and pig production decreases the risk of falling into poverty. The only agricultural commodity for which we do not find any statistically significant effects of increased values is coffee, which is not surprising in view of the results in Table 5.

Looking at measures of the household assets, the value of productive assets per capita only marginally increases the change of escaping poverty and the risk of falling into poverty is likewise only marginally reduced. Moreover, access to clean water does not seem to affect the chance of escaping poverty or the risk of falling into poverty. As for the presence of a permanent road, these results do not mean that productive assets and access to clean water are not determining the poverty status of households but rather that these assets are strong predictors of the chronic poor and the never poor, as seen from the chance of being non-poor in both periods relative to being poor in both periods and from the marginal effects of changes in these assets on the probability of these states.

Finally, living in households without electricity in the dwelling significantly increases the probability of being poor in both periods as the chance of escaping poverty is 40 percent lower in households without electricity relative to households that have electricity. Still, the presence of electricity is mainly separating the chronic poor from the never poor.

Overall, the model appears to provide a reasonable basis for discriminating between the chronic poor and the more affluent individuals that are never poor. But we do not really capture the movements in and out of poverty from 2002 to 2004. There are many possible explanations for this result, one of the more obvious being that such movements are dominated by idiosyncratic events

many of which occur in the interim period from 2002 to 2004. Another possible explanation may be that ethnicity has an impact on the relative risk ratios whereby the poverty transitions should be modeled using two different regressions for the Kinh-Hoa majority and the ethnic minorities. We explore this possibility in the remainder of the section.

Table 8 reports regression results, given as the percentage change in the relative risk, for the two sub-samples consisting of the Kinh-Hoa majority (2471 observations in 1001 primary sampling units) and the ethnic minority group (486 observations in 229 PSUs). The marginal effects of changes in the explanatory variables are given in Tables 9 and 10. The results for the Kinh-Hoa majority are based on the same baseline individual as we used for the full sample in Table 7 while we move the ethnic minority baseline individual to the Northern Uplands. The change in the baseline for the minority group is driven by the fact that the Northern Uplands has the largest minority population. The estimated probabilities at which we evaluate the marginal effects of changes in the explanatory variables are very different for the Kinh-Hoa and the ethnic minorities (see the last rows of Tables 9 and 10, respectively). We believe the baseline probabilities are reasonable for each of the two sub-populations, but it should be noted that these differences makes direct comparisons of the magnitudes of the marginal effects across the sub-groups meaningless. The percentage changes in the relative risks are directly comparable, though.

		Kinh-Hoa							Ethnic minorities						
				Percei	ntage ch	n relative risk (DRR)									
	Pr(P -	- NP)	Pr(NF	Pr(NP – P) Pr		-NP)	Pr(P -	- NP)	Pr(NP	– P)	Pr(NP-	NP)			
	Pr(P	– P)	Pr(NP	– NP)	Pr(P	– P)	Pr(P	– P)	Pr(NP -	- NP)	Pr(P -	- P)			
Northern Uplands	-75.2	***	53.6		-75.1	***	-73.6		1076.6	**	-95.7	***			
Red River Delta	-49.5	*	31.1		-83.4	***	-87.3		-		-100.0	***			
North Central Coast	-75.5	***	36.9		-88.8	***	-81.5		46.4		-97.6	***			
South Central Coast	-3.6		53.3		-18.1		-98.0	***	449.3		-98.9	**			
Central Highlands	-47.9		-40.3		-74.3	***	-61.7		3088.5	***	-96.1	***			
South East	476.3	***	-48.5		829.1	***	-10.0		112.8		-86.6				
Household characteristics															
Female head	-1.8		-7.2		6.7		-61.8		-4.8		580.0				
Age of head	-0.7		-5.0	***	4.3	***	0.1		-0.2		5.7	**			
Age of head, squared	0.0		0.1		-0.1	*	-0.1		0.0		-0.3				
No spouse	67.0		98.4		2.5		-90.0	**	387.9		-98.9	***			
Share of children <sup>1</sup>	-20.1		56.9	***	-93.8	***	-19.8		115.7	*	-97.9	**			
Share of elderly <sup>1</sup>	-1.6		81.1	***	-92.4	***	-7.4		-20.4		-71.3				
Log of household size <sup>2</sup>	2.4		11.4	*	-23.6	***	-19.0	*	-11.9		-40.7	***			
Head's education and occupation															
Primary school	39.1		-50.6	**	103.2	***	51.4		49.1		0.4				
Low. secondary school	139.5	***	-46.2	**	236.5	***	130.6		-19.7		182.0	*			
High. secondary school	66.4		-47.7		311.2	***	-15.2		-		36.5				
Production	-13.9		-52.1	**	97.4	***	-		-91.7	*	996.4	***			
Sales	54.3		-77.2	**	401.7	***	182.4		-28.6		-21.7				
White-collar	132.9	*	-31.9		235.0	***	166.4		-51.6		1071.9	**			
Not working	-15.4		-47.7	*	53.2		251.5		430.5		87.7				
Agricultural production															
Rice	2.1		-2.9	***	5.7	***	7.0		-1.8		9.2	**			
Coffee	2.7	*	0.0		1.1		-20.2		-18.4		6.7				

 Table 8: Results of Multinomial Logistic Regression for the Rural Kinh-Hoa

 and Ethnic Minorities

			Kinh	-Hoa			Ethnic minorities							
			P	ercer	ntage cha	ו relative risk (DRR)								
	<u> Pr(P -</u>	<u> Pr(P – NP)</u>		<u> Pr(NP – P)</u>		<u>Pr(NP-NP)</u>		<u> Pr(P – NP)</u>		<u> Pr(NP – P)</u>		-NP)		
	Pr(P	– P)	Pr(NP -	- NP)	Pr(P -	- P)	Pr(P	– P)	Pr(NP	– NP)	Pr(P -	- P)		
Fruit	16.3	**	-0.7		20.5	***	13.6		-3.1		20.5			
Pigs	5.1	*	-6.1	***	12.2	***	10.3		-1.8		25.7	***		
Feathered animals	25.7	**	-5.0		32.1	***	-21.1		-19.6		53.2	***		
Aquaculture	2.6		-1.8		3.6	*	-3.9		-1.3		13.3	**		
Assets														
Productive assets	1.5	*	-0.6		4.8	***	3.0	**	-2.5		5.7	***		
No electricity	-56.6	***	41.9		-75.5	***	3.9		-4.6		-72.3	**		
Access to clean water	3.7		-21.5		140.9	***	-48.2		48.7		-12.8			
Commune characteristics														
Permanent road	21.4		9.5		-100.0	**	41.1		-81.2	**	-100.0	***		
Post office	-6.6		3.1		-8.4		0.6		-76.8	*	108.1			
Market	1.1		-33.9		21.9		-65.7	**	342.4	**	-60.2			
Lower secondary school	-22.7		-11.6		17.2		28.7		3.6		-39.9			
Agricultural extension center	-20.4		-8.9		-13.5		23.3		119.0		9.5			

*Notes:* \*, \*\* and \*\*\* indicates significance at the 10%, 5% and 1% level, respectively. <sup>1</sup>The effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The effect is evaluated at a change in the household size from five to six.

Source: Own calculations based on the VHLSS 2002 and 2004.

		P - P			P - NP			NP - P			NP - NP		
	ð i	$\tau_{il}/\partial\xi$	s.e.	ð 1	$\tau_{12}/\partial\xi$	s.e.	ð:	π,ϡδξ	s.e.	ð n	τ.1θξ	s.e.	X
Northern Uplands	29.1	***	8.3	-11.8	*	6.2	-0.6		3.3	-16.6	**	6.5	0
Red River Delta	26.0	***	7.1	4.2		7.0	-4.2		2.6	-26.0	***	5.1	0
North Central Coast	40.4	***	6.1	-8.0		5.8	-4.5	*	2.6	-27.8	***	5.3	0
South Central Coast	1.4		5.8	0.9		6.0	2.8		3.8	-5.2		6.0	0
Central Highlands	22.4	**	10.6	2.5		10.0	-5.6	**	2.8	-19.3	***	7.0	0
South East	-17.3	***	4.7	-2.0		6.7	-1.9		2.9	21.1	***	7.7	0
Household characteristics													
Female head	-0.4		5.7	-1.2		5.1	-0.3		2.9	1.9		5.6	0
Age of head	-0.3		0.2	-0.6	***	0.2	-0.2		0.1	1.1	***	0.2	44
No spouse	-4.7		5.6	8.6		7.0	4.7		5.2	-8.7		5.8	0
Share of children1	6.3	***	2.4	2.4		2.4	1.7		1.4	-10.4	***	2.4	0.4
Share of elderly1	4.4	*	2.4	5.8	**	2.7	2.4	*	1.3	-12.6	***	2.5	0
Log of household size2	2.4	**	1.1	4.2	***	1.2	-0.3		0.5	-6.3	***	1.1	ln(5)
Head's education and occupation													
Primary school	-7.3	**	3.4	-3.0		4.5	-2.9		1.9	13.1	***	4.3	0
Low. secondary school	-12.2	***	3.5	-0.8		5.0	-2.2		2.0	15.2	***	5.0	0
High. secondary school	-12.6	***	4.1	-10.5	*	6.2	-1.4		3.0	24.5	***	6.7	0
Production	-5.5		3.6	-10.8	***	3.9	-2.5		2.0	18.8	***	4.4	0
Sales	-13.4	***	4.1	-13.5	***	5.0	-4.9	*	2.6	31.8	***	5.4	0
White-collar	-12.3	***	4.3	-1.8		6.3	-0.7		3.1	14.7	**	7.3	0
Not working	-2.8		5.2	-8.0		5.6	-2.6		2.2	13.4		5.9	0
Agricultural production													
Rice	-0.7	***	0.2	-0.3		0.2	0.0		0.1	1.0	**	0.2	6.2
Coffee	-0.3		0.3	0.4		0.3	0.0		0.2	-0.1		0.5	0
Fruit	-2.8	**	1.3	0.4		0.7	0.4	*	0.2	2.1	***	0.8	0.1
Pigs	-1.4	***	0.5	-0.5		0.4	-0.1		0.2	2.0	***	0.4	2.5
Feathered animals	-4.2	***	1.6	0.8		1.4	0.2		0.4	3.2	**	1.3	0.5
Aquaculture	-0.5		0.3	0.1		0.2	-0.1		0.1	0.5	**	0.2	0
Assets and shocks													

Table 9: Estimated Marginal Effects for the Baseline Individual Rural Kinh-Hoa

		P - P			P - NP			NP - P			NP - NP		
	ð	$\pi_{il} \partial \xi$	s.e.	ð:	π,,/∂ξ	s.e.		∂π <sub>ι</sub> ∮δξ	s.e.	ð:	π.↓∂ξ	s.e.	)
Productive assets	-0.6	***	0.2	-0.4	**	0.2	0.1	**	0.1	0.8	***	0.2	8
No electricity	23.7	***	5.8	-1.9		5.4	-2.1		2.4	-19.7	***	5.0	0
Access to clean water	-8.3	***	3.2	-11.2	***	3.3	1.1		2.1	18.4	***	4.0	0
Commune characteristics													
Permanent road	-5.9	**	2.7	-3.9		3.5	2.4		2.1	7.4	*	3.8	0
Post office	1.3		3.7	-0.2		3.9	0.0		1.8	-1.1		4.0	0
Market	-1.5		3.3	-1.9		3.8	-2.1		1.7	5.4		4.3	0
Lower secondary school	-0.2		3.7	-6.9	**	3.3	0.2		2.0	6.8		4.5	0
Agricultural extension center	3.2		3.2	-2.4		3.4	-0.7		1.9	0.0		3.9	0
Pij	20.6			29.5			8.3			41.7			

Notes: Probabilities are given as percentages and marginal effects as percentage points. \*, \*\*, and \*\*\* indicates significance at the 10%, 5% and 1% level of significance, respectively. Marginal effects of categorical variables are computed as the change in the estimated probability when the indicator variable changes from 0 to 1. The base values of the regressors are given in the last column. <sup>1</sup>The marginal effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The marginal effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The marginal effect is evaluated for a change in the household size from five to six.

Source: Own calculations based on the VHLSS 2002 and 2004.

Change in Household's Poverty Status in Rural Vietnam 2002-2004

		P - P			P - N	Р	NP -	NP - NP				
	ð 1	τ,1/∂ξ	s.e.	дπ	τ,1θξ	s.e.	$\partial \pi_{a} \partial \xi$	s.e.	∂π	τ.1 <i>θ</i> ξ	s.e.	x
Northern Uplands	43.3	***	11.3	3.3		19.8	4.7	5.2	-51.4	***	19.1	1
Red River Delta	38.9	***	12.6	-18.4	*	9.9	-8.6	5.8	-11.9	*	7.2	0
North Central Coast	35.9	***	13.6	-16.3		10.8	-8.1	5.6	-11.4	*	6.9	0
South Central Coast	42.0	***	11.7	-22.6	***	8.1	-7.8	5.6	-11.7		7.2	0
Central Highlands	17.5		21.9	-11.6		15.3	5.4	12.7	-11.3		7.0	0
South East	13.0		34.2	2.6		32.5	-5.6	6.0	-9.9		7.1	0
Household characteristics												
Female head	-28.3		20.5	-19.0	**	8.4	19.1	20.0	28.2		34.5	0
Age of head	-0.6		0.5	-0.2		0.3	0.4	0.4	0.5		0.4	44
No spouse	39.0	***	10.5	-19.4	**	7.8	-7.9	5.2	-11.7		7.2	0
Share of children1	8.0		6.0	-1.8		5.7	1.2	3.0	-7.5		5.4	0.4
Share of elderly1	5.0		7.5	0.3		6.6	-3.4	4.6	-1.9		4.5	0
Log of household size2	9.4	***	3.0	-1.0		2.3	-4.2	2.8	-4.2		2.6	ln(5)
Head's education and occupation												( )
Primary school	-7.9		9.3	7.1		8.0	2.5	5.9	-1.6		5.5	0
Low. secondary school	-21.7	**	10.4	9.7		9.7	3.4	5.8	8.7		7.4	0
High. secondary school	4.8		20.3	-1.9		16.4	-8.6	5.8	5.7		10.8	0
Production	-27.2		20.9	-23.4	***	8.2	-4.6	6.4	55.2	***	19.2	0
Sales	-14.9		19.1	25.1		20.3	-5.1	5.4	-5.1		11.6	0
White-collar	-37.8	***	13.6	-3.1		14.1	7.3	12.6	33.6		20.7	0
Not working	-33.3	**	13.1	9.9		22.8	26.2	22.4	-2.9		8.8	0
Agricultural production												
Rice	-1.8	*	1.0	0.8		0.8	0.3	0.3	0.7		0.5	6.2
Coffee	3.2		2.5	-4.0		3.1	-0.7	1.2	1.5		0.9	0
Fruit	-3.7		3.3	1.5		2.1	0.8	1.4	1.4		1.6	0.1
Pigs	-3.8	**	1.9	0.7		1.3	1.2	1.0	1.9	*	1.1	2.5
Feathered animals	-0.7		4.0	-5.9	*	3.6	1.7	1.5	4.9	**	2.6	0.5
Aquaculture	-0.9		1.9	-1.3		2.1	0.8	0.7	1.3		0.9	0
Assets												

Table 10: Estimated Marginal Effects for the Baseline Individual Rural Ethnic Minorities

	P - P	P - P		P	NP -	Р	NP - NP		
	$\partial \pi_{u} / \partial \xi$	s.e.	$\partial \pi_d \partial \xi$	s.e.	$\partial \pi_{\mathcal{J}} \partial \xi$	s.e.	$\partial \pi_{i} \partial \xi$	s.e.	x
Productive assets	-0.9 ***	0.3	0.3	0.2	0.1	0.2	0.5	0.3	8
No electricity	9.2	8.5	4.9	6.6	-6.0	5.1	-8.1	5.7	0
Access to clean water	6.4	14.0	-9.9	10.2	3.8	7.8	-0.3	6.1	0
Commune characteristics									
Permanent road	-14.7	9.2	1.0	7.5	-4.5	4.9	18.3	9.3	0
Post office	-4.4	9.9	-1.7 **	8.2	-4.8	4.6	10.9	9.0	0
Market	10.7	12.8	-13.9	5.9	9.5	11.2	-6.3	4.7	0
Lower secondary school	0.7	8.7	7.1	8.0	-3.2	3.4	-4.7	4.4	0
Agricultural extension center	-8.8	9.6	0.9	7.8	8.8	7.6	-0.9	4.3	0
Pij	56.0		23.4		8.6		11.9		

Notes: Probabilities are given as percentages and marginal effects as percentage points. \*, \*\*, and \*\*\* indicates significance at the 10%, 5% and 1% level of significance, respectively. Marginal effects of categorical variables are computed as the change in the estimated probability when the indicator variable changes from 0 to 1. The base values of the regressors are given in the last column. <sup>1</sup>The marginal effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The marginal effect is evaluated for a change in the share of children/elderly of 0.2 corresponding to an additional child/elderly in a family of five. <sup>2</sup>The marginal effect is evaluated for a change in the household size from five to six.

Source: Own calculations based on the VHLSS 2002 and 2004.

The regression results for the Kinh-Hoa majority in Table 8 reveal that the full sample results are dominated by the Kinh-Hoa sub-sample. The differences between the estimated changes in risks are small and, individually, all parameter estimates for the Kinh-Hoa sub-sample in Table 8 are within the 95% confidence bounds for the full sample estimates. Further, the baseline individual has almost identical estimated probabilities of being in each of the poverty states in the full sample and the Kinh-Hoa sub-sample regressions and the marginal effects of changes in the explanatory variables are thus very similar. We do find a couple of small changes in that an increased value of rice production in 2002 does not significantly increase the change of escaping poverty while an increased value of coffee production in 2002 is now marginally significant at the 10% level and the estimated order of magnitude corresponds to an increase in the value of rice production. Although this result should not be overstated (the result does not carry over to significant marginal effects of increases in coffee production for the baseline individual) it is in accordance with the findings in Nimi *et al.* (2003, 2004) who find rather large effects of increased coffee production on the chance of escaping poverty in the 1990s.

The results for the ethnic minorities are hampered by the small size of the minority sample. In some cases we find rather extreme changes in the relative risk ratios, such as positive changes in excess of 1000 percent or negative changes very close to -100 percent. In addition, the parameter estimates are in general very imprecisely determined in the ethnic minority sub-sample. A few results are worth noting, though. First of all, the strong effect of location found for the Kinh-Hoa, whereby living in the southern parts of Vietnam implies better chances of escaping poverty, does not seem to include the ethnic minorities. Even the minorities living in the proximity of Ho Chi Minh City in the South East Region do not seem to have much better chances of escaping poverty than the minority people living in other regions. This is in accordance with the results discussed in Section 2. Second, although often insignificant, the estimates of the impact of education and occupation point towards high returns to education and outside job opportunities for the minorities, as noted in Section 2. In particular, occupation within production or a white-collar job is a strong indication that the individuals in the household are non-poor in both periods. Further, while we find no significant effects on the chance of escaping poverty or the risk of falling into poverty, there is a strong effect on the chance of being non-poor in both periods relative to being poor in both periods of increases in the value of agricultural production and these chances appear somewhat stronger for the ethnic minority group compared to the Kinh-Hoa. In particular, production of pigs and feathered animals seem to have high returns in terms of the probability of being non-poor. While the direction of causality must be examined further, the high return to livestock production points towards a possible agricultural policy in which the ethnic minorities are encouraged to engage in and increase their production of pigs and birds.

Finally, for the ethnic minorities we find some significant effects of local infrastructure. Specifically, the presence of a permanent road or a post office reduces the vulnerability of the non-poor as the risk of falling into poverty is 80 percent lower when the commune has a permanent road or a post office. In, contrast we find a perverse effect of local markets as the presence of a market in 2002 seem to decrease the chance of escaping poverty and at the same time increase the risk of falling into poverty. One explanation for this result may be that the model is not including information about markets established between 2002 and 2004 in combination with the strong support for establishing local markets. The data shows that while only 32 percent of the ethnic minorities live in communes having a market in 2002 this share increases to 43 percent in 2004. If markets do lead to rapid poverty reduction this increase of 11 percentage points may well be large enough to generate the result. This highlights one of the serious limitations of our modeling strategy.

In sum, the split into the sub-samples consisting of the Kinh-Hoa majority and the ethnic minorities does not significantly improve the model's ability to depict systematic determinants of

movements in and out of poverty from 2002 to 2004. But we find that the full sample results are essentially describing the Kinh-Hoa majority. Unfortunately it is difficult to assess with any certainty if the systematic determinants have different impacts on the minorities' poverty transitions because of the small sample of minority households.

#### 4. Conclusion

In this chapter we have sought to establish an extended poverty profile for Vietnam that not only characterizes the poor and non-poor at a given point in time but, instead, describes poverty transitions over a short period of time. The panel data component of the rural parts of the nationally representative household surveys VHLSS 2002 and 2004, consisting of 3146 rural households, provides the basis for the analysis. The panel data shows a large outflow from poverty from 2002 to 2004 (15.3 percent of the rural population) and a comparably smaller inflow to poverty (4.8 percent). The resulting decline in the rural poverty rate is substantial, but despite the large decline, 19.6 percent of the rural population was poor in both 2002 and 2004.

In our analysis of the gross flows in and out of poverty from 2002 to 2004 we describe the variations in the flows using location, education, occupation, agricultural production, household assets, local infrastructure and ethnicity as the main determinants, much in line with classical poverty profiles. We describe the variations in the flows using both cross tabulations and an econometric model that analyzes the joint impact of the determinants. While we have difficulties in precisely describing the group of people who escape poverty and, in particular, the relatively small group of people who fall into poverty in the period, the model has good predictive power when it comes to discriminating between the chronic poor and the never poor. In some sense we show that it may be more meaningful to describe three groups of individuals: the very (or chronic) poor, the vulnerable; being the individuals close to the poverty line—both just above and below, and the more affluent who are never poor.

The descriptive analysis and the econometric model both lead to results that, although they are not surprising, illustrate the significance of the poverty determinants in a dynamic perspective. Most importantly, as the rural poverty rate has declined over time it has become increasingly clear that Vietnam is facing a serious problem with respect to including the ethnic minorities in the economic and social success story. Our estimates show that the ethnic minority people have a much smaller chance of escaping poverty than the Kinh and Hoa majority even when differences in location, education and occupation are taken into account. While our results do not provide direct guidance in terms of policies for reducing poverty among the ethnic minorities, a couple of findings are worth noting. First, the presence of a permanent road appears to lower the risk of falling into poverty for the non-poor minorities. This indicates that local infrastructure investments may be important for the vulnerable ethnic minority people just above the poverty line. Second, livestock production seems to have a high return in the sense that those minority households who have a relatively high monetary value of pig and bird production in 2002 are likely to be non-poor in both 2002 and 2004. Although this result must be analyzed in more detail before it can be used as a suggestion for an agricultural policy, it indicates that livestock production may be a profitable agricultural activity for (some of) the ethnic minorities.

Among the other determinants we find that both education and non-farm jobs are, individually, reducing the probability of being poor. Primary schooling reduces the risk of falling into poverty while secondary schooling both reduces the risk of falling into poverty and increases the chance of escaping poverty. This supports the continued effort for increasing the share of people who completes secondary schooling. Regarding non-farm employment we find that some occupations (production and sales) have protected against falling into poverty while a large fraction of the initially poor white-

collar workers have escaped poverty. However, we emphasize that non-farm work is not the only path out of poverty for the rural population. A larger agricultural production (or higher prices) is also associated with a better chance for escaping poverty. The only exceptions are coffee production and aquaculture. Hence, policies targeted at increasing the productivity in agriculture will probably still be poverty reducing in the future.

#### Acknowledgements

The authors would like to thank Nguyen Thang and Bob Baulch for fruitful discussions and comments on various drafts of this chapter.

#### References

- Baulch, B., and N. McCulloch. 1998. "Being Poor and Becoming Poor: Poverty Status and Poverty Transitions in Rural Pakistan." IDS Working Paper No. 79. http://www.ntd.co.uk/idsbookshop/details.asp?id=451
- Carter, M. R., and J. May. 2001. "One kind of freedom: poverty dynamics in post-apartheid South Africa." *World Development* 29(12): 1987-2006.
- Dercon, S., and P. Krishnan. 2000. "Vulnerability, seasonality and poverty in Ethiopia." *Journal of Development Studies* 36(6): 25–53.
- Freije, S. 2000. "Income, positional and poverty dynamics in Venezuela." NIP-LACEA Conference. Rio de Janeiro. http://www.lacea.org/meeting2000/SamuelFreije.pdf.
- Glewwe, P., and G. Hall. 1995. "Who is Most Vulnerable to Macroeconomic Shocks? Hypothesis Tests Using Panel Data from Peru." LSMS Working Paper No. 117. The World Bank.
- Glewwe, P., Gragnolati, M., and H. Zaman. 2002. "Who Gained from Vietnam's Boom in the 1990s?" *Economic Development and Cultural Change* 50(4): 773-92.
- Herrera, J. 2001. "Poverty Dynamics in Peru, 1997-1999." Working paper DT/2001-09. DIAL, Paris. http://www.dial.prd.fr/dial\_publications/PDF/Doc\_travail/2001-09.pdf
- Herrera, J., and F. Roubaud. 2005. "Urban Poverty Dynamics in Peru and Madagascar, 1997-99: A Panel Data Analysis." *International Planning Studies* 10(1): 21–48.
- Jalan, J., and M. Ravallion. 1998. "Transient Poverty in Postreform Rural China." *Journal of Comparative Economics* 26: 338–357.
- Justino, P., and J. Litchfeld. 2003. "Poverty Dynamics in Rural Vietnam: Winners and Losers during Reform. PRUS Working Paper No.10. University of Sussex. http://www.sussex.ac.uk/Units/PRU/wps/wp10.pdf
- Kedir, A. M., and A. McKay. 2005. "Chronic Poverty in Urban Ethiopia: Panel Data Evidence." International Planning Studies 10(1): 49–67.
- Lawson, D., McKay, A., and J. Okidi. 2005. "Poverty Persistence and Transitions in Uganda: A Combined Qualitative and Quantitative Analysis." http://aae.wisc.edu/seminars/papers/2005%20Fall%20papers/devecon/lawson.10.27.pdf
- Litchfield, J. and P. Justino. 2004. "Welfare in Vietnam During the 1990s: Poverty, Inequality and Poverty Dynamics." *Journal of the Asia Pacific Economy* 9(2): 145-169.
- McCulloch, N., and M. Calandrino. 2003. "Vulnerability and Chronic Poverty in Rural Sichuan." World Development 31(3): 611-628.
- Mu, R. and D. van de Walle. 2007. "Rural Roads and Poor Area Development in Vietnam." Policy Research Working Paper 4340. The World Bank.
- Nimii, Y., P. Vasudeva-Dutta, P and A. Winters. 2003. "Trade Liberalization and Poverty Dynamics in Vietnam." PRUS Working Paper No. 17. University of Sussex.

- Nimii, Y., P. Vasudeva-Dutta and A. Winters. 2004. "Storm in a Rice Bowl: Rice Reform and Poverty in Vietnam in the 1990s." *Journal of the Asia Pacific Economy* 9(2): 170-190.
- Okrasa. W. 1999. "Who avoids and Who escapes from Poverty during Transition: Evidence from Polish Panel 1993-1996.". Project RPO 681-21. The World Bank. http://wbln0018.worldbank.org/Research/workpapers.nsf/f7ff86a8cbbd411a852566db005f7b60/d d6911d066c4565b8525682c005b0d3b/\$FILE/wps2221.prn.pdf
- SRV [Socialist Republic of Vietnam]. 2003. National Education for All (EFA) Action Plan 2003 2015. Hanoi, Vietnam.
- VASS [Vietnamese Academy of Social Sciences]. 2006. Poverty Update Report 2006. Hanoi.
- World Bank. 1999. Vietnam Development Report 2000: Attacking Poverty. Hanoi.

World Bank. 2003. Vietnam Development Report 2004: Poverty. Hanoi.

## 7

### A Look at Parameter Constancy and Poverty Predictions in Poverty Mapping Models

#### Henrik Hansen, Pham Huong Giang and Vu Hoang Dat

#### 1. Introduction

Poverty is not uniformly distributed across areas, neither globally nor nationally. In almost every country, communes with high incidences of poverty coexist with communes of relatively high prosperity. Locating the poor areas in a country is of importance for policymakers if they are to make growth and development inclusive. However, the perception of poverty and well-being often vary across local areas resulting in heterogeneous treatment of the poor within a country. To overcome this problem the central government must ensure that poverty is measured in a coherent way across locations. This is not always an easy task. Poverty maps can assist the policymakers by locating poor areas within a country using a consistent poverty measure.

Poverty maps provide a description of the spatial distribution of poverty at detailed administrative levels such as provinces, districts and in some cases even communes. Partly because of the policy importance and partly because of a strong support from the World Bank, poverty mapping has become an important and popular tool in investigations and discussions of social, environmental and economic problems in many developing countries. Over a period of less than 10 years about 25 developing countries have computed detailed poverty maps using the technique of small area estimation in a form initially suggested in Hentschel, Lanjouw, Lanjouw, and Poggi (2000) and further developed in Elbers, Lanjouw and Lanjouw (2000, 2002 and 2003).<sup>117</sup>

The small area estimation idea is both simple and powerful.<sup>118</sup> Sample surveys with detailed information about income and expenditure that can be used for poverty statistics are too small—in terms sample size—to be useful in estimation of poverty statistics at geographical levels below broad regions of a country. Census data, on the other hand, while having sufficient sample sizes typically do not include detailed information about household income and expenditure permitting estimation of income or expenditure based poverty statistics. The small area estimation technique combines the information from a survey and a census by using a model based prediction of, say, per capita expenditure for all households in a small area. These predictions can subsequently be used to estimate poverty and inequality statistics for the local area.

<sup>&</sup>lt;sup>117</sup> See Demombynes et al. (2002, 2003) and Henninger and Snel (2002) among many others.

<sup>&</sup>lt;sup>118</sup> See Ghosh and Rao (1994) and Rao (1999, 2003) for a general introduction to small area estimation.

The poverty map estimation techniques suggested by Hentschel *et al.* (2000) and Elbers *et al.* (2000, 2002 and 2003) is often described as a three stage procedure (Lanjouw, 2003). In the first stage variables that are common in both a survey and a census are identified. The main issue in this step is to ensure that the common variables are defined and measured in exactly the same way in both data sets. The second stage consists of formulation and estimation of a consumption model using the variables identified in the first step. Estimation of the consumption model parameters is, naturally, based on the survey sample as the census does not include the consumption data. In the third stage the estimated consumption model parameters are used jointly with the census data to predict real per capita consumption expenditure for each household member represented in the census and, based on these predictions, poverty and inequality statistics can be computed for small areas as the census now has imputed data on consumption expenditure.

As seen from the brief description, the small area estimation approach is built on the key assumption that the survey and the census represent the same underlying population. An important implication is that the survey and the census should cover (exactly) the same point in time. This requirement is often difficult to meet. Table 1 lists some 24 countries, which have estimated poverty maps.<sup>119</sup> The second column in the table reports the year of the census while the third column reports the year of the survey used for estimation of the consumption function parameters. Most poverty mappings using recent census data have either no, or a very short, time span between the census and the survey. However, moving back in time the differences between census and survey year increases to 4-5 years and in one case even to 9 years.

	•		
Country	Census year	Survey year	Difference (years)
Burkina Faso	1985	1994	-9
Ecuador	1990	1994	-4
Panama	1990	1997	-7
Uganda	1991	1992	-1
Madagascar	1993	1993	0
Peru	1993	1997	-4
Guatemala	1994	1998/99	-4/5
Morocco	1994	1998	-4
Lao PDR	1995	1997/98	-2/3
Nicaragua	1995	1998	-3
South Africa	1996	1995	1
Mozambique	1997	1996/97	1/0
Malawi	1998	1997/98	1/0
Cambodia	1998	1999	-1
Azerbaijan	1999	2002	-3
Kenya	1999	1997	2
Vietnam	1999	1997/98	2/1
Thailand	2000	2000	0
Philippines	2000	2000	0
Bangladesh	2001	2000	1
Bulgaria	2001	2001	0
Jamaica	2001	2001	0
Sri Lanka	2001	2002	-1

#### **Table 1: Countries with Poverty Mappings**

Source: Henninger and Snel (2002) and documentation available at <u>www.worldbank.org/poverty</u> - under topic "Poverty Analysis", section "Mapping Poverty".

<sup>&</sup>lt;sup>119</sup> Information about the poverty map for each of the countries can be found in documents on the World Bank's PovertyNet website for poverty mapping – www.worldbank.org/poverty - under topic "Poverty Analysis", section "Mapping Poverty".

Having a time span of several years between the census and the survey, the strict assumption of describing the same underlying population is difficult to uphold. Yet, it is still possible to apply the small area estimation technique if one makes the weaker assumption that the underlying superpopulation is unchanged such that the parameters in the consumption model are constant over time. In this case it is still meaningful to predict consumption levels and poverty statistics in the census year.

Even though the assumption of parameter constancy is crucial to poverty mappings that do not have census and survey data for the same year it has never been investigated systematically. Accordingly, we set out to look at the importance of the assumption of constant parameters in the consumption model using data from three household living standard measurement surveys and one census in Vietnam. As seen from Table 1, Vietnam already has a poverty map. Minot and Baulch (2004) predicted poverty rates at the provincial level by combining the Vietnam Living Standards Survey from 1998 with a three percent sample of the Population and Housing Census 1999 while Minot, Baulch, and Epprecht (2003) predicted poverty rates at the district and commune level using the same survey and consumption model and a 33 percent sample of the Population and Housing Census 1999. These poverty maps have been widely used in the policy debate in Vietnam (see *e.g.*, Inter-Ministerial Poverty Mapping Task Force, 2003 and World Bank and others, 2003).

In this chapter, we pose the simple question; what if the Vietnamese poverty map had been based on the VHLSS 2002 or the VHLSS 2004 survey in the estimation of the parameters of the consumption model instead of the VLSS 1998 survey? Even though the time span between the 1999 census and the VHLSS 2002 survey is larger than between the VLSS 1998 survey and the census the difference of 3 years is still not in any way out of the ordinary (see Table 1). In addition to comparing parameters and poverty maps based on the three surveys we also, as a by product, compare poverty predictions based on the Hentschel *et al.* (2000) method and predictions based on the more general method developed by Elbers *et al.* (2003). This comparison is natural as Minot and Baulch (2004) used the first method in their provincial level poverty mapping, while we use the second method.<sup>120</sup>

Vietnam is an interesting case to study because the economic growth and poverty reduction has been extraordinary since the early 1990s (see, *inter alia*, Glewwe, 2004 and World Bank, 2003). Four living standard measurement surveys have been conducted in Vietnam since 1993 and the poverty statistics illustrate the unique development. Table 2 reports the poverty incidences for Vietnam as a whole and by region. The figures in Table 2 clearly show that in terms of poverty and, hence, per capita consumption expenditure one cannot speak of a constant population underlying the four surveys. This does not, however, exclude the possibility of a constant super-population.

<sup>&</sup>lt;sup>120</sup> Researchers at the World Bank have developed the program PovMap, which makes it straightforward to compute poverty mappings based on the Elbers *et al.* (2003) method. The new poverty mappings estimated in this chapter are all computed using PovMap 2.0. The program can be downloaded from http://iresearch.worldbank.org/PovMap/
		Poverty Incidence (percent)			Average annual change (percentage point)		
	1993	1998	2002	2004	1993 -1998	1998 -2002	2002 -2004
North East	86.1	62.0	38.4	29.4	-4.8	-5.9	-4.5
North West	81.0	73.4	68.0	58.6	-1.5	-1.4	-4.7
Red River Delta	62.7	29.3	22.4	12.1	-6.7	-1.7	-5.1
North Central Coast	74.5	48.1	43.9	31.9	-5.3	-1.1	-6.0
South Central Coast	47.2	34.5	25.2	19.0	-2.5	-2.3	-3.1
Central Highlands	70.0	52.4	51.8	33.1	-3.5	-0.2	-9.3
South East	37.0	12.2	10.6	5.4	-5.0	-0.4	-2.6
Mekong River Delta	47.1	36.9	23.4	15.9	-2.0	-3.4	-3.8
Vietnam	58.1	37.4	28.9	19.5	-4.1	-2.1	-4.7

## Table 2: Poverty Incidence based on Four Household Living Standard Measurement Surveys.

Source: VLSS 1993, VLSS 1998 VHLSS 2002 and VHLSS 2004.

The rapid decline in poverty in Vietnam accentuates the question one may have regarding constancy of the consumption function parameters. On one hand, if the parameters are not constant, the poverty mapping exercise may be of limited interest in the case of Vietnam as a mapping based on data from 1998 and 1999 may have little relevant information for policy makers today. On the other hand, if the parameters are constant there is a good opportunity for increasing the precision of the poverty map by including the much larger survey VHLSS 2002 in the consumption function estimation.<sup>121</sup> Furthermore, if the VHLSS 2002 combined with the census results in reasonable poverty estimates, there may be a way to update the poverty map to more recent years as the VHLSS 2002 and the VHLSS 2004 have a panel component of about 4,000 households. Following the suggestion in Hoogeveen, Emwanu and Okwi (2003) this panel can be used to update the poverty map. However, we will not cover this aspect in the present chapter.

An interesting aspect of parameter constancy in poverty mappings is that the relevant criterion is not parameter constancy *per se* but constancy of the predictions of per capita expenditure. Moreover, since the poverty and inequality statistics are non-linear functions of per capita expenditure, the relevant criterion is constancy of the non-linear transformation of the consumption model predictions. This means that constancy of the parameters in the consumption model is a sufficient but not a necessary condition for constancy of poverty mapping statistics. In effect we need to evaluate the predictive accuracy of the poverty predictions. Hence we combine the poverty mapping technique with forecast evaluation statistics.

The chapter is organized as follows. The next section presents the poverty mapping technique and the methods we use to evaluate differences in predictions. The data is presented in Section 3, while the consumption model regression results are given in Section 4. In section 5 the predictions based on combining survey and census data are compared. Finally, Section 6 offers concluding remarks.

## 2. The Estimation Technique and Prediction Comparison Statistics

In this section we first describe the formulation and estimation of the per capita expenditure model. Subsequently, estimation of the welfare measure, which in this chapter is limited to the poverty

<sup>&</sup>lt;sup>121</sup> The VLSS 1998 surveyed 6,000 households while the VHLSS 2002 surveyed 30,000 households.

incidence, is described. Most of the section is based on Elbers, Lanjouw and Lanjouw (2002, 2003); henceforth denoted ELL.

## 2.1. The Expenditure Regression Model

We consider the logarithm of per capita expenditure for household *h* in stratum *s*, ln  $y_{sh}$ . The conditional expectation of ln  $y_{sh}$  is assumed to be linear at the in the set of *k* observable variables that are common in the survey and the census with constant parameters at some stratum level.<sup>122</sup>

$$\ln y_{sh} = x_{sh}\beta_s + u_{sh},\tag{1}$$

where  $x_{sh}$  is a 1 x k vector of regressors and  $\beta_s$  is a k x 1, stratum specific, coefficient vector.

Given the degrees of freedom in most survey samples, weighted least squares using the sampling weights is often considered sufficient in expenditure regressions. However, the poverty statistics are nonlinear functions of per capita expenditure, which is why modeling the variance is also important. Therefore an explicit model for the disturbances is formulated.<sup>123</sup> The variance of the household specific disturbance term is modeled as a parametric function of a subset of the covariates (denoted  $z_{sh}$ ). To ensure non-negative estimates, ELL suggests a logistic transformation, and this suggestion is followed in most applications.

Letting  $A^* = (1.05) \max(\hat{u}_{sh}^2)$  in which  $\hat{u}_{sk}$  are the residuals from the consumption model (1), a variance regression model can be formulated as

$$\ln\left(\frac{\hat{u}_{sh}^2}{A^* - \hat{u}_{sh}^2}\right) = z_{sh}\alpha_s + r_{sh},\tag{2}$$

where  $z_{sh}$  is the 1 x k' vector of variance model regressors and  $\alpha_s$  is a k' x 1 coefficient vector. The variance regression model is used to form an estimate of the variance for the household level disturbance

$$\hat{\sigma}_{sh}^{2} = A^{*} \left( \frac{\exp(z_{sh}\hat{\alpha}_{s})}{1 + \exp(z_{sh}\hat{\alpha}_{s})} + \frac{\hat{\sigma}_{r}^{2}}{2} \frac{\exp(z_{sh}\hat{\alpha}_{s})(1 - \exp(z_{sh}\hat{\alpha}_{s}))}{(1 + \exp(z_{sh}\hat{\alpha}_{s}))^{3}} \right)$$
(3)

where  $\hat{\sigma}_r^2$  is the estimated variance of the variance model error term,  $r_{sh}$ .

The variance estimates are subsequently used in combination with the sampling weights in a second round estimation of the consumption model (1) to obtain FGLS estimates of  $\beta_s$ .

#### 2.2. The Small Area Poverty Incidence Estimator

Given the consumption model parameters one can predict (or impute) log per capita expenditure for each household in the census. But the predicted per capita expenditure is not the statistic of interest in the poverty mapping. The statistics of interest are, however, functions of the predictions. ELL describe a large class of poverty and inequality measures by the generic formulation of welfare measures,  $W(m_a, X_a, \beta_a, u_a)$ , where subscript *a* denotes the area covered by the measure (*e.g.*, a given province, district or commune), and  $m_a$  is an  $M_a$ -vector of household sizes used to convert the household unit measure (per capita expenditure at the household level) to an individual unit.  $X_a$  is the  $M_a \times k$  matrix of observable variables for the households in the area while  $u_a$  is an  $M_a$ -vector of the unknown

<sup>&</sup>lt;sup>122</sup> We follow Minot and Baulch (2004), and allow for different parameters across rural and urban areas. Elbers *et al.* (2002, 2003) split the model according to the eight strata in their survey

<sup>&</sup>lt;sup>123</sup> ELL considers both cluster correlation and heteroskedasticity. Here, we look at the provincial level poverty incidence, which is a much higher level of aggregation than the enumeration areas. Therefore, we omit specific treatment of cluster effects in the following.

disturbance terms for each household in the area. The vector  $\beta_n$  is the area specific per capita consumption model parameter vector that may be a combination of several stratum level parameters.

In this chapter the only welfare measure we consider is the poverty incidence. In terms of the welfare function the poverty incidence, with a poverty line w, is given as

$$W(m_a, X_a, \beta_a, u_a) = \sum_{i=1}^{M_a} \omega_i m_i \mathbf{1}(y_i < w) = \sum_{i=1}^{M_a} \omega_i m_i \mathbf{1}(x_i \beta_a + u_i - \ln w < 0)$$
(4)

where the summation is over all households in the chosen area,  $\mathbf{1}(\cdot)$  is the indicator function and  $\omega_i$ ,  $m_i$ ,  $x_i$ , and  $u_i$  are the household specific sampling weights, household sizes, observables and disturbances, respectively.<sup>124</sup>

Obviously, neither the levels of consumption nor the disturbances are observed in the census. However, if the disturbance terms, *u<sub>i</sub>*, are normally distributed, the poverty rates can be estimated by

$$\hat{\mu} = \hat{E}(W \mid m_a, X_a, \hat{\theta}_a) = \sum_{i=1}^{M_a} \omega_i m_i \Phi\left(\frac{\ln w - x_i \hat{\beta}_a}{\hat{\sigma}_i}\right)$$
(5)

where  $x_i \hat{\beta}_a$  is the estimated per capita log-expenditure for household *i*, computed from (1),  $\hat{\sigma}_i$  is the household specific standard deviation of the disturbance, computed from (3), using the census variables, and  $\theta_a = (\beta'_a, \alpha'_a, \sigma^2_r)'$  is the (k+k'+1)-vector of all parameters in the consumption model.

As seen, the estimated poverty incidence for an area is simply the weighted average of the predicted probability of an individual in a household being below the poverty line.

Next, obtaining a measure of precision of the estimated poverty incidence is also important in the poverty map estimation. Hentschel *et al.* (2000) and Minot and Baulch (2004) derive analytic expressions for the variance of the estimated poverty incidence given an assumption of homoskedastic, within-stratum, disturbances. However, allowing for heteroskedasticity (and cluster correlation) makes the analytical solutions intractable. This why ELL use simulation estimation of the variance. Such a simulation estimation can be done in several ways. Here, we describe the simplest (and most direct) simulation algorithm, which can be expressed as a three stage procedure:<sup>125</sup>

1. Two parameters vectors,  $\tilde{\beta}_a$ ,  $\tilde{\alpha}_a$ , are drawn randomly from the sampling distributions of the estimated parameters  $\tilde{\beta}_a \sim N(\hat{\beta}_a, \hat{\Sigma}_{\beta})$  and  $\tilde{\alpha}_a \sim N(\hat{\alpha}_a, \hat{\Sigma}_{\alpha})$ .

2. An *M*<sub>a</sub>-vector of normally distributed disturbances is drawn conditional on the household specific variance, which is computed using  $\tilde{\alpha}_a$  from step 1.

3. An estimate of the poverty incidence,  $\tilde{W}_a$ , is computed using equation (4) with observables from the census, the parameters from step 1 and the disturbances from step 2.

When steps 1 to 3 are repeated R times, the mean and variance of the poverty incidence can be estimated as

$$\tilde{\mu}_a = \frac{1}{R} \sum_{j=1}^R \tilde{W}_{aj},\tag{6}$$

and

$$V(\tilde{\mu}_{a}) = \frac{1}{R} \sum_{j=1}^{R} (\tilde{W}_{a} - \tilde{\mu})^{2} + \frac{1}{R} \operatorname{Var}(\hat{\mu}_{a}).$$
(7)

<sup>&</sup>lt;sup>124</sup> When a full census is available the sampling weights reduce to  $1/M_a$ .

<sup>&</sup>lt;sup>125</sup> See Elbers et al. (2000, 2002 and 2003) for details of other simulation algorithms.

The last term in the variance estimate is the simulation error variance, which can be made arbitrarily small by using many repetitions. In this study we use R = 1000, which appears more than sufficient to make the simulation error variance negligible.

As noted in Minot and Baulch (2004) there is an additional variance term when the poverty incidence estimate is based on a sample of the census instead of the full census. This variance term is estimated by assuming the estimated probability of being poor,  $\Phi\{(\ln w - x_i \hat{\beta}_a) / \hat{\sigma}_i\}$ , is fixed for each household and, subsequently, estimating the census sampling variance of the mean using the sampling weights, the stratum information and the enumeration clusters from the census sample. Denoting this census variance  $V_C(\hat{\mu}_a)$ , the resulting standard error of the estimated poverty incidence is estimated as

$$\operatorname{se}(\tilde{\mu}_a) = \sqrt{V(\tilde{\mu}_a) + V_C(\hat{\mu}_a)}.$$
(8)

The estimated local area poverty incidences,  $\tilde{\mu}_a$ , and their standard errors, se( $\tilde{\mu}_a$ ), are the statistics of interest in poverty mapping studies. However, in the present study we are also interested in the underlying parameters as we wish to test for constancy of these parameters across different surveys.

#### 2.3. Tests of Parameter Constancy and Prediction Equality

In comparisons of poverty incidence predictions based on different surveys, the prediction formula (5) clearly shows that parameter constancy (*i.e.*, constancy of  $\theta_a$ ) is a sufficient, but not a necessary condition for constancy of the predicted poverty incidence. Hence, we need to test both constancy of the parameters and equality of the predictions.

In testing parameter constancy and prediction equality one may use a number of different tests. However, one important distinction is that tests of parameter constancy only involve the survey data while tests of prediction equality involve both the surveys and the census. Ultimately, it is the tests of prediction equality for the census data that are of interest but much can be learned by analyzing the expenditure regression models in some detail using the survey data.

Turning to the hypotheses and the test statistics, consider the case in which two surveys are available for the expenditure model both of which have the same variables in common with the census, but no overlap in terms of households (that is, there is no panel component in the two surveys). Let  $\theta_s^{(1)}$  and  $\theta_s^{(2)}$  be the stratum level parameter vectors for the two surveys and let the hat denote the estimates. The first test of constancy is whether the stratum level parameters, or some subset thereof, are equal

$$H_0: \theta_s^{(1)} = \theta_s^{(2)}$$
 for all s,

or

$$H_0^{\beta}: \beta_s^{(1)} = \beta_s^{(2)} \text{ and } H_0^{\alpha}: \alpha_s^{(1)} = \alpha_s^{(2)} \text{ for all } s$$

These hypotheses can be tested using Chow tests.

Next, for the poverty incidence predictions the main problem is that we do not have the true poverty incidences. Hence, it is not possible to compare the prediction errors. Yet, given two surveys, two sets of local area predictions of poverty incidences can be generated and these may be compared under a null hypothesis that the two sets of predictions are random samples of matched pairs drawn from the same distribution. Based on this simple idea we look at several tests for equality of the predictions.

Two finite sample tests that have proven powerful in prediction comparisons are the sign test and Wilcoxon's signed rank test. A third test is the standard student-*t* test. Finally, the *t*-test can be extended to a regression-based test that provides more information. We briefly present the four tests below.

(9)

Let  $\tilde{\mu}_a^{(1)}$  and  $\tilde{\mu}_a^{(2)}$  be predictions of poverty incidences for area *a*, and let *d*<sup>*a*</sup> be the difference between the predictions. One possible null-hypothesis of equal accuracy of the two predictions is

 $H_0: E(\tilde{\mu}_a^{(1)}) = E(\tilde{\mu}_a^{(2)})$  for all *a*,

or

$$H_0: E(\tilde{\mu}_a^{(1)} - \tilde{\mu}_a^{(2)}) = E(d_a) = 0 \text{ for all } a,$$
(10)

which is just a statement that the population mean of the prediction difference is zero. Another possible null-hypothesis is a zero median difference:
(11)

$$\operatorname{med}(\mu_a^{(1)} - \mu_a^{(2)}) = \operatorname{med}(d_a) = 0.$$

The null-hypothesis of a zero mean difference is only identical with the hypothesis of a zero median difference if the distribution of the differences is symmetric. The latter hypothesis can be tested using the sign test and Wilcoxon's signed rank test.

The sign test can be based on the statistic

$$S_1 = \sum_{a=1}^{N} \mathbf{1}(d_a > 0), \tag{12}$$

where  $\mathbf{1}(d_a > 0)$  is the indicator function for a positive difference and *N* is the number of areas. The sign test is testing the hypothesis

$$H_0^{(1)}: \frac{E(S_1)}{N} = \frac{1}{2} \quad \text{vs} \quad H_1^{(1)}: \frac{E(S_1)}{N} \neq \frac{1}{2}$$
(13)

and the *p*-value of the test can be computed from the cumulative binomial distribution  $Bn(S_1, 0.5, N)$ . If the hypothesis is not rejected, we cannot reject that the median of the difference between the two predictions is zero.

Next, Wilcoxon's signed rank test is based on the statistic

$$S_2 = \sum_{a}^{N} \mathbf{1}(d_a > 0) \operatorname{rank}\left(\left|d_a\right|\right)$$
(14)

and a studentized version of the test has a limiting normal distribution under the null hypothesis

$$\frac{S_2 - N(N+1)/4}{\sqrt{N(N+1)(2N+1)/24}} \stackrel{a}{\sim} N(0,1)$$
(15)

As for the sign test, we cannot reject a zero median of the difference between the two predictions if the test statistic in (15) is insignificant.

Third, if the distribution of prediction differences is normal, the most powerful test for comparing the predictions is the student-*t* statistic. The *t*-statistic is computed as

$$t = \sqrt{N} \frac{d}{\sqrt{\operatorname{Var}(d_a)}} \sim t(N-1), \tag{16}$$

where d is the sample average of the prediction differences and  $Var(d_a)$  is the sample variance.

The three tests above do not use information about the size of the predictions, as the test statistics only include the prediction differences. However, there may be systematic differences in the local area poverty incidence predictions, which are related to the size of the poverty incidence. For instance, the prediction differences may be positive at low levels of poverty and negative at high levels of poverty. To assess if this is so we use a regression approach and test if the prediction differences are A Look at Parameter Constancy and Poverty Predictions in Poverty Mapping Models

systematically related to the level of the predictions. Specifically, consider a regression of prediction 1 on prediction 2:

$$\tilde{\mu}_a^1 = \gamma_0 + \gamma_1 \tilde{\mu}_a^2 + \mathcal{E}_a. \tag{17}$$

By subtracting prediction 2 from both sides we obtain the relationship between the prediction differences and the prediction level

$$d_a \equiv \tilde{\mu}_a^{(1)} - \tilde{\mu}_a^{(2)} = \gamma_0 + (\gamma_1 - 1)\tilde{\mu}_a^{(2)} + \mathcal{E}_a.$$
<sup>(18)</sup>

The null hypothesis of equality of the underlying distribution of poverty incidences is now given by  $(\gamma_{1}-1) = 0$  and  $\gamma_{0} = 0$ . The *t*-statistic of  $(\gamma_{1}-1) = 0$  tests if the correlation between the predictions is 1 and the joint test of significance of both parameters is a test of unit correlation and equal means (an extension of the *t*-test above). If  $(\gamma_{1}-1) < 0$ , there is a systematic tendency for having positive differences at low levels of poverty and negative differences at high levels of poverty. This tendency is the reverse when  $(\gamma_{1}-1) > 0$ .

If the null hypotheses cannot be rejected for any of the four tests—assuming the normality assumption for the *t*-test is valid—we conclude that the predictions are from the same distribution and, hence that the predictions are equally good estimates of the unknown true local area poverty incidences.

## 3. The Data

We use three Vietnamese household living standard surveys; VLSS 1998, VHLSS 2002, and VHLSS 2004, and a 3 percent sample of the 1999 Population and Housing Census to estimate three provincial level poverty maps of Vietnam.

The VLSS 1998 is described in more detail in World Bank (2001). The survey covered 6,000 households, selected using a stratified random sample. Sample weights are provided with the survey data. The VHLSS 2002 and the VHLSS 2004 covered 30,000 and 9.000 households, respectively. About 4,000 of these households are interviewed in both surveys implying that the observations are not independent. This issue will be ignored in the empirical analysis as the 2004 survey is mainly included to highlight possible systematic differences in the predictions based on the 2002 and 2004 surveys. The two VHLSS surveys also used stratified samples and sampling weights are included with the survey data.<sup>126</sup> In all three surveys the possible cluster correlation due to enumeration units is at the lowest administrative level (the commune in rural areas and the district/ward in urban areas). Hence, in the following, provincial level data and above are considered to be independent observations.

The Population and Housing Census 1999 refer to the situation as of April 1, 1999. The 3 percent sample was selected by the GSO using a stratified random sample of 5,287 enumeration units and 534,139 households. The sample was designed to be representative at the provincial level.

<sup>&</sup>lt;sup>126</sup> The appendix to this book provides a comprehensive description of the two VHLSS surveys.

Variable name(s) in Minot	
and Baulch (2004)	Description of variable
hhsize	Household size (number of people)
pelderly	Proportion of elderly people (age over 60) in household
pchild	Proportion of children (aged under 15) in household
pfemale	Proportion of females in household
ledchd_1 to 6	Highest level of education completed by head (less than primary school, primary school, lower secondary school, upper secondary school technical or vocation training, college diploma or university degree)
ledcsp_0	Dummy for no spouse
ledcsp1 to 6	Highest level of education completed by spouse (less than primary, primary school, lower secondary school, upper secondary school technical or vocation training, college diploma or university degree)
ethnic	Indicator of ethnic minority head (not Kinh or Hoa)
loccup_1 to 7	Occupation of head over last 12 months (political leader or manager, professional or technical worker, clerk or service
	worker, agriculture non-farm enterprises, unskilled worker, not-working)
Ihouse_1 to 3	Type of house (permanent; semi-permanent or wooden frame, simple)
htypla1 to 2	House type interacted with the logarithm of the living area (m <sup>2</sup> )
electric	Household with electricity
lwater_1 to 3	Main source of drinking water (private or public tap, rainwater and wells, rivers and lakes)
Itoilet_1 to 3	Type of toilet (flush, latrine/other, none)
tv	Indicator of TV ownership
radio	Indicator of radio ownership
reg7_1 to 7	Regional indicators (7 regions)

Table 3: Household Characteristics Common to	o the Census
and the Household Surveys	

Source: Minot and Baulch (2004).

The common variables in the census and the VLSS 1998 were identified by Minot and Baluch (2004). Table 3 presents information about the 17 variables they include in the poverty mapping. The variables cover household demographics in the form of household size and composition; ethnicity (Kinh/Hoa or other); education of both household head and spouse; occupation of household head; housing in the form of type of house; water source; energy source (electricity) and type of toilet. Finally, there are seven regional indicators. Today, Vietnam has eight regions and 64 provinces. But in 1999 there were only seven regions and 61 provinces. When using the 2002 and 2004 surveys we remap the 64 provinces to 61 and the eight regions to the seven regions.

Tables 4 and 5 report the weighted means of the variables we are using in the expenditure regression for the census and each of the three surveys by rural and urban region. The weighted means are estimates of the population means in the years of the surveys and the census. So, ideally, the weighted means should be equal if they are describing the same underlying population. In the tables the superscript *a* in indicates that the mean of the specific variable in the survey is not statistically different from the census mean of the variable at the 5 percent level of significance; the superscript *b* indicates that the mean of the variable is not statistically different from the VLSS 1998 mean (only applicable for VHLLS 2002 and 2004); finally superscript *c* indicates that the mean in the VHLSS 2004 is not statistically different from the mean in VHLSS 2002.

	Census 1999	VLSS 1998	VHLSS 2002	VHLSS 2004
Log real per capita income		7.561	7.671	7.811
Household composition				
Household size	4.559	5.549	5.178	5.055
Proportion of elderly people (>60)	0.110	0.097	0.087	0.092 <sup>bc</sup>
Proportion of children (<15)	0.340	0.345 <sup>a</sup>	0.306	0.276
Proportion of females	0.521	0.513	0.503	0.493
Ethnic minority head Education of household head	0.126	0.180	0.151 <sup>b</sup>	0.160 <sup>bc</sup>
Primary school	0.483	0.237	0.253 <sup>b</sup>	0.263 <sup>c</sup>
Lower secondary school	0.085	0.277	0.281 <sup>b</sup>	0.299 <sup>b</sup>
Upper secondary school	0.061	0.041	0.071	0.060ª
Technical or vocation training	0.051	0.048 <sup>a</sup>	0.036	0.036 <sup>c</sup>
College diploma/University degree Education of spouse	0.014	0.011ª	0.013 <sup>ab</sup>	0.014 <sup>abc</sup>
No spouse	0.211	0.137	0.141 <sup>b</sup>	0.150 <sup>bc</sup>
Primary school	0.397	0.179	0.220	0.218 <sup>c</sup>
Lower secondary school	0.055	0.204	0.236	0.245 <sup>c</sup>
Upper secondary school	0.037	0.031ª	0.045	0.042 <sup>c</sup>
Technical or vocation training	0.027	0.027ª	0.021 <sup>b</sup>	0.019 <sup>bc</sup>
College diploma/University degree	0.008	0.005 <sup>a</sup>	0.010 <sup>a</sup>	0.011 <sup>ac</sup>
Occupation of household head				
Leaders in al fields	0.006	0.016	0.015 <sup>b</sup>	0.022 <sup>b</sup>
Professionals in all fields	0.026	0.027 <sup>a</sup>	0.017	0.019 <sup>c</sup>
Clerk or service workers	0.027	0.047	0.018	0.023 <sup>a</sup>
Skilled workers in agriculture	0.679	0.701 <sup>a</sup>	0.624	0.592
Skilled workers	0.054	0.072	0.080 <sup>b</sup>	0.088 <sup>bc</sup>
Unskilled workers and army force Housing	0.021	0.062	0.139	0.138 <sup>c</sup>
Permanent house	0.088	0.088 <sup>a</sup>	0.124	0.147
Semi-permanent house	0.645	0.618ª	0.608 <sup>b</sup>	0.618 <sup>bc</sup>
Permanent house x log(area)	0.339	0.341ª	0.501	0.616
Semi-permanent house x log(area)	2.385	2.352 <sup>a</sup>	2.407 <sup>ab</sup>	2.485 <sup>b</sup>
Electricity	0.724	0.705 <sup>a</sup>	0.830	0.910
Private or public tap water	0.020	0.019 <sup>a</sup>	0.057	0.063 <sup>c</sup>
Rain water and wells	0.715	0.679 <sup>a</sup>	0.662 <sup>b</sup>	0.685 <sup>b</sup>
Flush toilet	0.043	0.037ª	0.115	0.164
Latrine/other toilet	0.771	0.740 <sup>a</sup>	0.696 <sup>b</sup>	0.672
TV ownership	0.461	0.514	0.640	0.762
Radio ownership Regional indicators	0.416	0.469	0.302	0.271
Red River Delta	0.227	0.188	0.198 <sup>b</sup>	0.198 <sup>bc</sup>
North Central Coast	0.150	0.163ª	0.153 <sup>ab</sup>	0.151 <sup>abc</sup>
South Central Coast	0.082	0.096	0.099 <sup>b</sup>	0.101 <sup>bc</sup>
Central Highlands	0.037	0.047	0.056	0.055 <sup>c</sup>
South East	0.105	0.077	0.077 <sup>b</sup>	0.079 <sup>bc</sup>
Mekong River Delta	0.220	0.225ª	0.229 <sup>ab</sup>	0.226 <sup>abc</sup>

Table 4: Means of Regressors in the Census and the Three Surveys. Rural area

*Notes:* <sup>a</sup> indicates an insignificant difference between the survey mean and the Census mean at a 5 percent level. <sup>b</sup> indicates an insignificant difference between the VLHSS 2002 and 2004 mean and the mean in the VLSS 1998 survey. <sup>c</sup> indicates an insignificant difference between the mean in VHLSS 2004 and 2002.

Source: Own calculations based on VLSS 1998, VHLSS 2002, VHLSS 2004 and Population and Housing Census 1999.

	Census 1999	VLSS 1998	VHLSS 2002	VHLSS 2004
Log real per capita income		8.293	8.457	8.580
Household composition				
Household size	4.361	5.221	4.938 <sup>b</sup>	4.889 <sup>c</sup>
Proportion of elderly people (>60)	0.082	0.117	0.086 <sup>a</sup>	0.094 <sup>c</sup>
Proportion of children (<15)	0.280	0.244	0.243 <sup>b</sup>	0.217
Proportion of females	0.520	0.526 <sup>a</sup>	0.508	0.508 <sup>c</sup>
Ethnic minority head	0.033	0.010	0.040 <sup>a</sup>	0.030 <sup>ac</sup>
Education of household head				
Primary school	0.302	0.208	0.216 <sup>b</sup>	0.212 <sup>bc</sup>
Lower secondary school	0.125	0.256	0.212	0.247 <sup>b</sup>
Upper secondary school	0.120	0.086	0.124 <sup>a</sup>	0.114 <sup>ac</sup>
Technical or vocation training	0.158	0.114	0.117 <sup>b</sup>	0.086 <sup>b</sup>
College diploma/University degree	0.099	0.086 <sup>a</sup>	0.111 <sup>ab</sup>	0.121 <sup>ac</sup>
Education of spouse				
No spouse	0.321	0.207	0.208 <sup>b</sup>	0.204 <sup>bc</sup>
Primary school	0.237	0.163	0.189 <sup>b</sup>	0.178 <sup>bc</sup>
Lower secondary school	0.088	0.211	0.171	0.216 <sup>b</sup>
Upper secondary school	0.084	0.056	0.110	0.087
Technical or vocation training	0.094	0.090 <sup>a</sup>	0.093 <sup>ab</sup>	0.080 <sup>abc</sup>
College diploma/University degree	0.056	0.056 <sup>a</sup>	0.072 <sup>b</sup>	0.076 <sup>c</sup>
Occupation of household head				
Leaders in al fields	0.017	0.032	0.025 <sup>b</sup>	0.030 <sup>bc</sup>
Professionals in all fields	0.112	0.100ª	0.094 <sup>b</sup>	0.106 <sup>abc</sup>
Clerk or service workers	0.134	0.264	0.073	0.081°
Skilled workers in agriculture	0.144	0.149ª	0.157 <sup>ab</sup>	0.134 <sup>abc</sup>
Skilled workers	0.178	0.190ª	0.150	0.148°
Unskilled workers and army force	0.080	0.064	0.251	0.249°
Housing				
Permanent house	0.265	0.361	0.317 <sup>b</sup>	0.382 <sup>b</sup>
Semi-permanent house	0.620	0.500	0.563	0.530 <sup>bc</sup>
Permanent house x log(area)	1.058	1.417	1.344 <sup>b</sup>	1.634 <sup>b</sup>
Semi-permanent house x log(area)	2,299	1.832	2.215ª	2.120°
Electricity	0.958	0.982	0.986 <sup>b</sup>	0.992 <sup>bc</sup>
Private or public tap water	0.470	0.578	0.535 <sup>b</sup>	0.554 <sup>bc</sup>
Rain water and wells	0.452	0.316	0.372 <sup>b</sup>	0.352 <sup>bc</sup>
Flush toilet	0.558	0.615 <sup>a</sup>	0.681 <sup>b</sup>	0.760
Latrine/other toilet	0.357	0.257	0.252 <sup>b</sup>	0.183
TV ownership	0.764	0.822	0.888	0.932
Radio ownership	0.559	0.599 <sup>a</sup>	0.404	0.404°
Regional indicators				
Red River Delta	0.195	0.224 <sup>a</sup>	0.173 <sup>ab</sup>	0.174 <sup>abc</sup>
North Central Coast	0.072	0.053ª	0.073 <sup>ab</sup>	0.066 <sup>abc</sup>
South Central Coast	0.094	0.148 <sup>a</sup>	0.126 <sup>b</sup>	0.121 <sup>bc</sup>
Central Highlands	0.036		0.065	0.0610
South East	0,326	0.301ª	0.284 <sup>b</sup>	0.310 <sup>b</sup>
Mekong River Delta	0.147	0.181ª	0.158 <sup>ab</sup>	0.160 <sup>abc</sup>

Table 5: Means of Regressors in the Census and the Three Surveys. Urban Area

Notes: a indicates an insignificant difference between the survey mean and the Census mean at a 5 percent level.

<sup>b</sup> indicates an *in*significant difference between the VLHSS 2002 and 2004 mean and the mean in the VLSS 1998 survey. <sup>c</sup> indicates an *in*significant difference between the mean in VHLSS 2004 and 2002.

Source: Own calculations based on VLSS 1998 VHLSS 2002, VHLSS 2004 and Population and Housing Census 1999.

It is apparent that it is very unlikely if the three surveys are describing the 1999 population, as most of the survey-based means are statistically different from the census means. Some of these differences are probably of minor importance, yet, there are others for which the differences are somewhat worrying. In

particular, it is clear that the size of households is changing rapidly when looking at the average household size in the three surveys. Even so, the average household size is consistently larger in the three surveys compared to the census. The prime reason for the difference is probably the exclusion of many single person households in the sampling frame for the surveys (specifically student homes). This guess is supported by the large difference between the fraction of households without a spouse in the census and in the surveys. In the rural area the share of households without a spouse is underestimated by 6-7 percentage points, while it is as much as 10 percentage points in the urban area. In general, one finds that the household demographics (size and composition) are moving towards smaller households with fewer children. This change may have a significant impact on the poverty predictions because the welfare measure is per capita consumption, *i.e.*, there is no use of equivalence scales.

Another disturbing difference is in the level of education of both the household head and the spouse. In particular, there appear to be differences in the definition of completed lower secondary schooling. According to the census almost half of the household heads have completed primary schooling (48 percent) and only 8.5 percent have completed lower secondary schooling in the rural area. But the corresponding shares are about 25 percent for primary schooling and 28-30 percent for lower secondary schooling in the surveys. While the figures are somewhat lower for the spouses the pattern is similar, and the pattern can also be found in the urban area. Hence, while the shares of the populations that have completed either primary or lower secondary schooling are roughly the same, the distinction between the two levels vary significantly between the census and the surveys. Within the surveys, there is a fairly similar distribution of educational levels, in particular when comparing the VHLSS 2002 and the VHLSS 2004.

The occupational distributions also differ, mainly in the rural area. For the VHLSS 2002 and 2004 the reason is probably the movement of workers out of agriculture and into wage labor as unskilled workers.

Looking at the indicators for housing there is a good agreement between the census and the VLSS 1998 in the rural area, but not in the urban area. In general one finds that the quality of housing has improved from the VLSS 1998 and the census to the later surveys in 2002 and 2004. This is in accordance with the economic development in Vietnam, so these differences are not surprising. The same goes for TV and radio ownership. The share of households with a TV is increasing over the years while the share of households with a radio is decreasing. The later probably reflects that a TV is a good substitute for a radio.

Overall, the comparison of the means of the regressors in the census and the three surveys does not lead to preference of one household survey over the two others. None of the surveys match the census, so they do not represent the same population. Yet the super-population may still be the same, in the sense that the changes in the regressors are systematically related to changes in consumption. We examine this hypothesis in the next section.

## 4. Comparison of Regression Results for Three Household Surveys

Following Minot and Baulch (2004) we formulate the consumption model allowing for different parameters in rural and urban regions. The dependent variable is the log of per capita consumption in January 1998 prices. The consumption measure is comparable across the three surveys and the parameters of the model are initially estimated by weighted least squares using the sampling weights from the surveys as weights.

The consumption regressions for the rural areas are given in Table 6 while the corresponding results for the urban areas are shown in Table 7. The fit in all regressions appears reasonable as the models explain about 50-60 percent of the total variation in log per capita consumption. Because we expect to have heteroskedastic residuals, we only report one specification test for the regression; the RESET test for neglected nonlinearity. As seen from Tables 6 and 7 the RESET test is insignificant at

conventional levels of significance, with the regression for the urban areas using the VHLSS 2002 as the exception. However, for that model the null hypothesis of linearity cannot be rejected at the 1 percent level of significance, so we conclude that the linearity assumption is generally valid.

Survey	VLSS 1998		VHLSS	2002	VHLSS 2004	
Regressors	Coefficien	t t-value	Coefficient	t-value	Coefficient	t-value
Household size	-0.155	-11.31	-0.117	-16.18	-0.110	-9.62
Household size squared	0.006	5.58	0.005	7.74	0.004	3.58
Proportion of elderly people (>60)	-0.138	-3.86	-0.134	-7.01	-0.122	-4.33
Proportion of children (<15)	-0.313	-8.63	-0.456	-24.71	-0.499	-18.76
Proportion of females	-0.123	-3.61	-0.068	-3.96	-0.011	-0.38
Ethnic minority head	-0.077	-1.85	-0.170	-10.43	-0.225	-11.02
Primary school	0.057	3.37	0.056	5.79	0.065	4.61
Lower secondary school	0.087	4.44	0.072	6.69	0.132	7.92
Upper secondary school	0.089	3.32	0.108	7.00	0.125	5.38
Technical or vocation training	0.135	4.27	0.170	7.88	0.201	5.98
College diploma/University degree	0.255	4.86	0.219	6.58	0.230	5.73
No spouse	0.002	0.08	0.040	3.55	0.039	2.18
Primary school	0.008	0.41	0.025	2.51	0.019	1.27
l ower secondary school	0.016	0.72	0.037	3.22	0.022	1 29
Inner secondary school	0.011	0.29	0 100	5.38	0.046	1.65
Technical or vocation training	0.090	2 29	0.229	8 97	0 194	5.08
College diploma/Liniversity degree	0.000	2 71	0.321	8 17	0.270	6.07
Leaders in all fields	0.133	2 22	0.021	3 71	0.050	1 21
Professionals in all fields	0.100	3.00	0.000	3.52	0.000	2.60
Clerk or service workers	0.120	3.20	0.100	3.02	0.105	1.68
Skilled workers in agriculture	-0.026	-0.04	-0.025	-1 05	-0.050	-2 00
Skilled workers	0.020	1 71	0.023	1.00	0.033	0.63
Unskilled workers and army force	-0.070	-2.07	0.072	3.81	-0.010	-0.45
Pormanont house	-0.070	-4.00	-0.410	-5.79	-0.776	-0.40
Somi-pormonont bouso	-0.932	-4.20	-0.419	-0.71	-0.770	-5.90
Dermanont house x log(area)	-0.310	-5.01	-0.410	0.00	-0.439	762
Somi permenent beuge x log(area)	0.290	5.01	0.137	10 10	0.240	7.00 0.75
	0.120	0.67	0.139	10.10	0.155	0.70
Electricity Drivete er public ten weter	0.076	2.07	0.005	4.09	0.057	2.03
Private or public tap water	0.080	1.30	0.129	5.03	0.104	3.95
Rain water and wells	0.114	4.30	0.046	3.95	0.027	1.80
Flush tollet	0.271	5.02	0.252	14.22	0.208	9.58
Latrine/other tollet	0.058	2.82	0.007	0.55	0.010	0.62
I v ownersnip Dedie europeine	0.217	14.83	0.206	25.95	0.229	18.23
Radio ownersnip	0.103	7.06	0.093	11.07	0.097	8.43
Red River Delta	0.025	0.47	-0.065	-3.93	-0.057	-2.97
North Central Coast	0.046	0.74	-0.124	-7.48	-0.128	-6.58
South Central Coast	0.136	2.21	0.021	1.08	-0.014	-0.59
Central Highlands	0.163	2.08	-0.023	-1.00	0.020	0.64
South East	0.532	9.46	0.265	11.28	0.269	10.01
Mekong River Delta	0.301	5.12	0.223	12.74	0.149	7.05
Constant	7.763	96.83	7.828	237.21	7.915	157.40
RESET test (p-value)	0.877		0.104		0.816	
R-squared	0.541		0.543		0.528	
Observations	4269		22621		6938	

**Table 6: Consumption Regressions. Rural Area** 

Notes: The dependent variable is log real per capita expenditure in January 1998 prices. The regressions are based on weighted least squares using the survey specific sampling weights. The t-statistics are cluster robust. *Source:* Own calculations based on VLSS 1998, VHLSS 2002 and VHLSS 2004.

Survey	VLSS 19	VLSS 1998		VHLSS 2002		VHLSS 2004	
Regressors	Coefficie	ent t-value	Coefficie	Coefficient t-value		ent t-value	
Household size	-0.176	-11.06	-0.136	-8.91	-0.156	-6.76	
Household size squared	0.007	6.57	0.005	4.09	0.008	4.03	
Proportion of elderly people (>60)	-0.156	-2.39	-0.147	-3.29	-0.069	-1.21	
Proportion of children (<15)	-0.207	-3.07	-0.348	-9.04	-0.273	-5.20	
Proportion of females	0.046	0.57	-0.050	-1.24	-0.139	-2.76	
Ethnic minority head	0.024	0.25	-0.168	-3.18	-0.188	-3.14	
Primary school	0.069	1.94	0.031	1.08	0.075	2.36	
Lower secondary school	0.029	1.04	0.098	3.40	0.167	5.10	
Upper secondary school	0.135	3.09	0.155	4.67	0.211	5.13	
Technical or vocation training	0.160	3.48	0.190	5.78	0.235	5.34	
College diploma/University degree	0.181	3.61	0.281	6.90	0.333	6.90	
No spouse	0.003	0.08	0.058	2.01	0.055	1.58	
Primary school	0.072	2.19	0.029	1.01	-0.025	-0.72	
Lower secondary school	0.100	2.81	0.055	1.84	0.005	0.16	
Upper secondary school	0.176	2.47	0.114	3.15	0.034	0.75	
Technical or vocation training	0.129	3.18	0.114	3.28	0.127	2.90	
College diploma/University degree	0.168	2.86	0.244	5.80	0.190	3.95	
Leaders in all fields	0.225	2.94	0.072	1.10	0.155	2.48	
Professionals in all fields	0.051	1.07	0.053	1.83	0.019	0.44	
Clerk or service workers	0.028	0.73	0.043	1.40	-0.020	-0.51	
Skilled workers in agriculture	-0.017	-0.35	-0.071	-2.90	-0.033	-0.97	
Skilled workers	-0.003	-0.08	-0.023	-0.85	-0.034	-1.07	
Unskilled workers and army force	-0.153	-2.84	0.004	0.18	-0.006	-0.19	
Permanent house	-0.609	-3.89	-0.477	-4.49	-0.676	-4.85	
Semi-permanent house	-0.451	-4.04	-0.415	-4.74	-0.416	-3.58	
Permanent house x log(area)	0.222	5.86	0.198	7.85	0.221	6.94	
Semi-permanent house x log(area)	0.154	5.08	0.142	6.56	0.122	4.38	
Electricity	-0.008	-0.12	-0.053	-0.74	0.129	1.18	
Private or public tap water	0.244	6.17	0.192	5.76	0.140	3.46	
Rain water and wells	0.042	0.81	-0.014	-0.44	-0.032	-0.81	
Flush toilet	0.134	2.05	0.187	5.28	0.217	4.19	
Latrine/other toilet	0.012	0.21	-0.061	-1.82	-0.108	-2.20	
TV ownership	0.226	5.92	0.252	8.92	0.273	7.86	
Radio ownership	0.160	5.87	0.129	7.72	0.127	5.97	
Red River Delta	0.084	0.87	0.085	2.17	0.129	3.26	
North Central Coast	0.049	0.57	-0.129	-3.75	-0.115	-2.88	
South Central Coast	0.174	2.22	-0.004	-0.10	0.069	1.68	
Central Highlands			-0.091	-1.92	-0.022	-0.40	
South East	0.426	5.59	0.406	10.67	0.457	11.89	
Mekong River Delta	0.199	2.19	0.063	1.79	0.103	2.34	
Constant	8.023	64.12	8.199	76.15	8.132	51.08	
RESET test (p-value)	0.102		0.020		0.536		
R-squared	0.562		0.631		0.610		
Observations	1730		6909		2250		

 Table 7: Consumption Regressions. Urban Area

Notes: The dependent variable is log real per capita expenditure in January 1998 prices. The regressions are based on weighted least squares using the survey specific sampling weights. The *t*-statistics are cluster robust. *Source:* Own calculations based on VLSS 1998, VHLSS 2002 and VHLSS 2004.

The only difference in the formulation of the consumption model compared to the model in Minot and Baulch (2004) is that we include the square of the household size. Using a non-linear model for the impact of household size on consumption is fairly standard in reduced form consumption models and we find the squared household size to be statistically significant in all regressions. The total impact of household size on log per capita consumption is negative in all regressions in the typical range of household sizes, say, 1 to 10, but the marginal effect of an extra household member is decreasing fairly rapidly.

The regression coefficients have the expected signs, at least when they are statistically different from zero. Households with higher proportions of elderly people and children have lower per capita consumption, all other things equal; the level of consumption is generally increasing with the educational level of the household head and the spouse, and in the rural areas there is a clear effect of the occupation of the household head. The occupation effect is not well determined in the urban areas, though.

For the housing variables there is a positive effect of having a permanent and a semi-permanent house compared to a simple house.<sup>127</sup> Furthermore, in the rural areas households with electricity are better off compared to similar households without electricity. This effect is not significant in any of the regressions for the urban area, but this is due to the very small fraction of households without electricity in the urban areas (see Table 5). Having access to a private or public water tab is also associated with higher per capita consumption levels both in the rural and urban areas. Using rain water and wells (instead of rivers and lakes) is only associated with higher consumption in the rural areas, but this is not surprising. Finally, having a flush toilet is associated with higher consumption levels both in the rural and urban areas, but access to a latrine or other toilet is only significantly positive in the rural area in the VHLSS 1998 regression. In the urban areas we record a significantly negative effect in the VHLSS 2004 regression. However, as only 6 percent of the urban households do not have a toilet according to the VHLSS 2004 survey this small comparison group may well be better off, on average, than households with a latrine or other toilet.

Finally, ownership of the durable goods TV and radio are significantly associated with higher per capita levels of consumption in all three surveys. Households with TVs consume roughly 25 percent more than similar households without TVs in the rural areas, and the ratio is even higher in the urban areas. Radio ownership is associated with about 10 percent higher per capita consumption levels in the rural areas, and slightly more in urban areas.

Turning next to the variance models, we use the residuals from the consumption regressions above to estimate the household specific variances, as explained in Section 2. The regression model is given in equation (2) and the regression results are shown in Table 8. As for the consumption regressions the parameters of the variance model are estimated using weighted least squares.<sup>128</sup>

<sup>&</sup>lt;sup>127</sup> Here, it is important to note that the type of house is interacted with the living area. Hence the negative point estimates on the indicators for permanent and semi-permanent houses have no independent interpretation as they measure the difference in consumption between households living in a permanent/semi-permanent house and households living in a simple house when the living area is one square meter. This size of living area is, obviously, never observed. When the impact of the house type is evaluated at the average living areas the impact is positive.

<sup>&</sup>lt;sup>128</sup> We use one common variance model for both the urban and rural areas. The main reason for this is that the program PovMap works with joint regression models for all strata. As seen from the regression specification, we are effectively estimating different variance models for the rural and urban areas because we interact all regressors with the urban indicator variable. The consumption regressions in Tables 3 and 4 can also be presented as a single, joint, regression by interaction with the urban indicator variable. In fact, this is the way the model is formulated in the PovMap program.

Survey	VLSS 1998 VHLSS 2002		2002	VHLSS 2	2004	
Regressors	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
(xb) <sup>2</sup>	0.0242	3.30	0.0330	10.40	0.0364	6.40
(xb) <sup>2</sup> x Ethnic minority head, rural	0.0005	0.27	0.0041	4.63	0.0056	3.50
(xb) <sup>2</sup> x Household size, rural	-0.0009	-0.70	-0.0020	-3.86	-0.0027	-3.16
(xb) <sup>2</sup> x Household size sq., rural	0.0001	0.62	0.0002	3.80	0.0002	3.19
(xb) <sup>2</sup> x Household size, urban	0.0038	2.98	0.0016	2.08	0.0012	0.76
(xb) <sup>2</sup> x Household size sq., urban	-0.0003	-3.19	-0.0001	-2.15	-0.0002	-1.81
(xb) <sup>2</sup> x Urban	-0.4891	-2.13	-0.2475	-3.11	-0.2928	-1.81
(xb) x Urban	8.0209	2.11	3.7187	2.78	4.3342	1.57
Urban	-33.6070	-2.13	-14.3179	-2.55	-16.3280	-1.39
Constant	-6.0839	-14.20	-7.0745	-37.74	-6.6054	-18.75
RESET test (p-value)	0.859		0.895		0.007	
Heteroskedasticity <sup>a</sup> (p-value)	0.111		0.863		0.220	
Heteroskedasticity <sup>b</sup> (p-value)	0.081		0.548		0.011	
R-squared	0.011		0.010		0.007	
Observations	5999		29530		9188	

Table 8: Variance Regression. Joint for Rural and Urban Areas

*Note:* The dependent variable is the logistic transformation of the squared residuals from the consumption regressions as shown in equation (3). The regressions are based on weighted least squares using the survey specific sampling weights. (xb) are the fitted values from the consumption regressions in Table and Table .

<sup>a</sup> White's test for heteroskedasticity. <sup>b</sup> Breuch and Pagan's test for heteroskedasticity.

Source: Own calculations based on VLSS 1998, VHLSS 2002 and VHLSS 2004.

The main difficulty in the formulation of the variance model is choice of regressors. Since there is little guidance from economic theory, the specification of the model may end up being a data mining exercise. To avoid this problem we keep the model simple by selecting only a few regressors. The main regressor is the square of the fitted value from the consumption regressions. A positive association between the consumption level and the variance across households is a fairly well established empirical regularity. To allow for different variances in the rural and urban areas we also include the indicator for urban areas and its interaction with the fitted value, and the fitted value squared. In addition, we add the household size and its square, both interacted with the fitted value squared and the urban indicator. The final regressor is the indicator for ethnic minority headed households which we also interact with the square of the fitted value and, in this case, only for the rural area. The reason for omitting the ethnic minority indicator in the urban area is the very low percentage of urban ethnic minority households (see Table 5). In total, the variance model has 10 regressors. Given that all regressors, except for the urban area indicator, are interactions with the square of the fitted value the individual parameter estimates have little interest. They are mainly changing the association between the fitted consumption level and the variance.

The fits of the variance models are less impressive than the fits of the consumption models as they only explain about one percent of the total variation in the dependent variable (the logistic transformation of the squared residuals). This is not surprising or problematic. The main purpose of the regression is to extract systematic variation in the disturbance variance, leaving a homoskedastic residual, but the systematic variation need not be a large fraction of the total variation. To assess the assumption of homoskedastic residuals in the variance equation we report two tests for homoskedasticity; White's general test and the Breuch-Pagan test in which we include all regressions we cannot reject the null hypothesis of a constant residual variance. In the VLSS 2002 regression White's test does not reject the null, but it is rejected by the more specific Breuch-Pagan test. In this

regression the hypothesis of linearity is also rejected. Thus, the model specification for VHLSS 2004 could be improved. However, the main focus in the present analysis is on the VLSS 1998 and VHLSS 2002. Therefore we have not searched for more general models for the VHLSS 2004.

Having estimated models for each of the three surveys we continue by estimating pair-wise joint models, and test for equality of the regression parameters using the Chow-test. The differences between the parameter estimates in the three consumption models are reported in Tables A1 and A2 in the Appendix while the differences between the parameters in the variance models are given in Table A3. In Table 9 we summarize the results by presenting the results of the overall Chow-tests.

Comparing the VLSS 1998 and the VHLSS 2002 regressions for the rural areas we find that 14 of the 41 parameters are, individually, significantly different. Therefore, it is not surprising that the overall Chow-test rejects the null hypothesis of equal parameters at conventional levels of significance. Moreover, when looking at the slope parameters as one group and the regional dummies as another group we also firmly reject equality of the parameters in each of the two groups. Hence, both the household specific variables and the regional indicators have different parameters in the two regressions. Equality of the parameters in the VLSS 1998 and VHLSS 2004 regressions is also rejected. Overall equality of the parameters in the rural area VHLSS 2002 and VHLSS 2004 regressions is also clearly rejected although equality of the regional dummy parameters is only rejected marginally. In the urban area regressions we cannot reject the hypothesis of equal intercepts and region parameters for any of the three pairs. However, as seen from Table 9, there is only one pair of consumption regressions for which we cannot reject overall parameter constancy. This is for the VHLSS 2002 and VHLSS 2004 urban area regressions in which none of the individual parameters are significantly different and we also accept joint constancy of all parameters.

For the variance models we also reject parameter constancy for all three model pairs. This result is slightly more surprising as the differences between the individual parameters are generally insignificant at the 5 percent level.

In sum, we conclude that none of the consumption models can be considered as describing the same super-population as they do not have the same systematic features. Still, this does not imply that the model based poverty predictions for the census population are significantly different. The final step in the comparison of the poverty mapping models is therefore testing whether the poverty predictions are equal, in the sense of being random realizations from the same underlying distribution.

Difference between	1998 and 2002	1998 and 2004	2002 and 2004
Consumption equations			
Rural area			
Chow-test	0.000	0.000	0.000
Chow-test of slopes	0.000	0.000	0.005
Chow-test of regions and intercept	0.000	0.002	0.032
Urban area			
Chow-test	0.000	0.000	0.427
Chow-test of slopes	0.000	0.000	0.487
Chow-test of regions and intercept	0.074	0.144	0.934
Variance equation			
Chow test	0.000	0.000	0.000

Table 9: Chow-tests of Parameter Constancy in the Consumption and Variance Models

Source: Own calculations based on VLSS 1998, VHLSS 2002 and VHLSS 2004.

## 5. Predictions of Provincial Poverty Rates Based on the Census 1999

The poverty incidences are predicted at the provincial level. As mentioned in Section 3, Vietnam had 61 provinces in 1999. The average number of households in the provinces is 8756 in the census sample. The smallest sample (Ninh Thuan) has 7521 households while largest (Nam Dinh) has 10235 households. Hence, the number of observations is sufficient for provincial level predictions even though we only have 3 percent of the population in the census sample.

We use the official GSO poverty line from 1998 as all regressions are based on per capita consumption in January 1998 prices. Accordingly, a household member is poor if the per capita consumption is below 1,789,871 VND per year.<sup>129</sup> Using predicted per capita consumption levels from the Census and the GSO poverty line, provincial poverty incidences are estimated as described in Section 2. We estimate three sets of provincial level poverty incidences based on the regression results from each of the three surveys VLSS 1998, VHLSS 2002 and VHLSS 2004. In addition, the predicted poverty incidences from Minot and Baulch (2004, Table 7.4) are also included in the tests for prediction equality. Minot and Baulch assumed a constant disturbance variance in the rural and the urban areas, respectively, and consequently used the method given in Hentschel *et al.* (2000) to predict provincial level poverty incidences. That is, they used equation (5) with the added assumption that the household specific variance is constant within the rural and within the urban areas. By including the Minot and Baulch (2004) predictions we can test if the two estimators give statistically different predictions.<sup>130</sup>

	VLSS	1998	Minot and	d Baulch	VHLS	S 2002	VHLSS	2004
Province	P0	se	P0	se	P0	se	P0	se
Lai Chau	77.8	2.5	77.7	3.4	75.4	2.5	73.1	2.8
Ha Giang	72.2	3.0	72.2	3.6	70.2	2.8	67.4	3.0
Son La	72.0	2.5	71.4	3.4	69.7	2.6	67.1	2.8
Cao Bang	68.4	2.6	67.5	3.3	65.1	2.2	62.2	2.5
Lao Cai	65.4	3.1	65.2	3.6	62.8	3.1	60.3	3.3
Lang Son	62.5	2.3	61.7	3.2	60.3	2.0	57.6	2.2
Bac Kan	60.9	2.7	60.9	3.4	59.2	2.3	56.5	2.5
Hoa Binh	59.0	2.6	58.6	3.6	58.4	2.4	56.1	2.6
Tuyen Quang	58.5	3.0	58.3	3.8	56.3	2.9	53.0	3.1
Gia Lai	57.2	3.1	53.8	4.7	63.5	2.8	57.6	3.1
Kon Tum	56.1	2.9	52.2	4.3	64.3	2.6	59.2	2.9
Yen Bai	54.5	2.9	55.0	3.6	52.3	2.8	49.2	3.0
Quang Tri	53.4	2.6	52.0	3.4	60.0	2.2	54.9	2.5
Ninh Thuan	53.0	2.4	52.5	3.3	51.8	2.3	47.8	2.5
Quang Binh	49.8	2.6	49.1	4.0	58.1	2.2	51.8	2.4
Thua Thien - Hue	48.4	2.5	47.2	3.3	52.8	2.4	47.5	2.6
Nghe An	48.3	3.1	47.7	4.1	57.8	2.7	52.3	3.0
Quang Ngai	48.3	3.0	47.4	3.8	49.9	2.8	46.2	3.0
Bac Giang	46.7	2.8	47.0	4.7	42.9	2.5	37.3	2.7
Thanh Hoa	46.4	2.7	46.0	4.1	57.1	2.5	52.2	2.8
Ha Tinh	44.6	2.3	44.5	4.0	55.4	1.8	48.7	2.1

Table 10: Estimated Provincial Poverty Incidence, 1999 (Percent)

<sup>129</sup> The poverty line is based on the cost of basic needs approach. Annual consumption expenditure of 1,789,871 VND is assumed to ensure an intake of 2100 calories per day. See the World Bank (1999) for a detailed description of the derivation of the poverty line.

<sup>130</sup> The VLSS 1998 data used in Minot and Baulch is identical to our VLSS 1998 data and we are very grateful to the two authors for giving us access to their survey data.

	VL	SS 1998	Minot	and Baulch	VHI	SS 2002	VHL	SS 2004
Province	P0	se	P0	se	P0	se	P0	se
Vinh Phuc	43.9	2.2	44.2	4.7	38.4	1.7	32.5	1.9
Binh Thuan	43.8	2.3	43.5	3.3	45.2	2.0	40.2	2.2
Phu Tho	42.8	2.5	43.1	4.2	38.8	2.3	34.0	2.4
Soc Trang	41.9	2.0	42.4	2.9	41.6	2.0	42.4	2.3
Dac Lak	41.8	2.6	39.5	5.0	51.5	2.3	42.6	2.7
Phu Yen	41.7	2.7	41.6	3.5	43.7	2.3	39.2	2.6
Thai Nguyen	41.2	2.1	41.9	3.8	38.3	1.9	34.3	2.0
Tra Vinh	41.2	1.9	41.8	3.0	39.5	1.8	40.0	2.1
Quang Nam	41.1	2.8	40.8	3.6	44.1	2.5	40.1	2.7
Ha Tay	40.0	2.2	39.5	3.1	41.9	1.9	34.7	2.1
An Giang	39.7	2.1	40.6	2.7	34.3	1.8	33.7	1.9
Binh Dinh	39.4	2.0	39.1	3.2	40.4	1.7	35.6	1.9
Ninh Binh	38.6	2.1	38.5	2.9	41.2	1.7	34.9	1.9
Hung Yen	38.3	2.0	38.3	3.0	41.6	1.6	34.6	1.8
Dong Thap	38.3	1.9	39.1	2.8	33.8	1.6	32.6	1.8
Ha Nam	37.5	2.2	37.6	3.1	40.0	1.8	33.5	1.9
Bac Ninh	37.4	2.2	38.3	4.6	32.6	1.6	27.4	1.7
Kien Giang	37.2	2.2	38.0	2.8	35.9	2.0	35.7	2.1
Bac Lieu	37.1	1.8	37.7	2.7	35.5	1.6	35.7	2.0
Lam Dong	35.6	2.4	33.7	3.9	42.7	2.2	35.8	2.6
Nam Dinh	34.9	1.8	35.1	2.8	38.0	1.4	31.5	1.7
Quang Ninh	34.3	2.3	35.7	2.9	31.0	2.1	29.1	2.2
Can Tho	33.8	1.8	34.9	2.5	30.0	1.5	29.3	1.7
Ca Mau	33.5	1.9	34.5	2.7	32.1	1.6	32.3	1.9
Thai Binh	32.4	2.0	33.0	3.2	34.9	1.6	29.7	1.8
Vinh Long	31.9	1.9	33.0	2.7	27.7	1.7	26.5	1.9
Hai Duong	31.5	1.7	31.9	2.8	34.4	1.3	28.8	1.5
Ben Tre	31.5	2.1	32.5	2.8	27.9	1.7	26.8	1.9
Khanh Hoa	30.6	2.0	31.1	2.7	31.1	1.8	27.4	1.9
Long An	29.4	1.9	30.5	2.7	26.7	1.5	26.6	1.6
Hai Phong city	28.2	1.6	28.6	2.2	29.3	1.4	24.5	1.5
Tien Giang	26.4	1.8	27.6	2.6	23.5	1.4	23.1	1.6
Binh Phuoc	17.5	2.0	17.9	2.4	29.2	2.4	26.3	2.7
Ha Noi city	14.7	1.1	15.2	1.5	14.6	1.1	12.0	1.0
Da Nang	13.9	1.3	15.6	1.9	14.3	1.3	10.8	1.1
Tay Ninh	11.9	12	12.4	1.6	20.3	14	16.8	1.5
Dong Nai	10.2	12	11 1	14	17.4	16	14.7	17
Ba Bia-Vung Tau	8 1	0.8	9.0	1.4	12.8	1.0	10.5	1.7
Binh Duong	6.8	0.0	70	1.1	11.0	1.1	0.0	1.1
Ho Chi Minh city	3.0	0.0	1.5	0.7	3.2	0.4	2.0	0.3
	3.2	0.5	4.4	0.7	3.2	0.4	2.3	0.3
Spearman Rank Correlations	_							
Minot and Baulch	0.997							
VHLSS 2002	0.960		0.943					
VHLSS 2004	0.963		0.953		0.983			

Notes: The poverty incidences and their standard errors based on the VLSS 1998, VHLSS 2002 and VHLSS 2004 regressions are estimated from 1000 simulations using the program PovMap2.0.

Source: Minot and Baulch (2004, Table 7.6) and own calculations based on VLSS 1998, VHLSS 2002, VHLSS 2004 and the Population and Housing Census 1999.

The provincial level poverty incidences are presented in Table 10. The provinces are listed according to the incidence of poverty as predicted by the VLSS 1998 based estimates. In the table, 'P0' is the estimated poverty incidence while 'se' is the associated standard error, which is estimated using equation (12).

Browsing Table 10 it is clear that the Minot and Baulch estimates are very close to our VLSS 1998 estimates while the VHLSS 2002 and VHLSS 2004 based estimates are somewhat different. Yet, there is a good agreement regarding the ranking of the richest and poorest provinces, and this is reflected in the Spearman rank correlations shown in the bottom part of Table 10. The rank correlations are all above 0.95 and, in particular, the rank correlation between the Minot and Baulch and the VLSS 1998 estimates is so close to one that the difference in the rankings could be due to rounding errors. Yet, the sizes of the estimates differ slightly, in particular in provinces with low poverty incidence.

	Minot and Baulch	VHLSS 2002	VHLSS 2004
P0(VLSS 1998)			
Sign test	0.200	1.000	0.000
Wilcoxon signed-rank test	0.144	0.277	0.000
t-test	0.724	0.065	0.000
Normality test	0.000	0.001	0.243
Average difference	0.048	1.108	-2.473
P0(Minot and Baulch)			
Sign test		0.609	0.000
Wilcoxon signed-rank test		0.399	0.000
t-test		0.117	0.000
Normality test		0.001	0.226
Average difference		1.060	-2.521
P0(VHLSS 2002)			
Sign test			0.000
Wilcoxon signed-rank test			0.000
t-test			0.000
Normality test			0.569
Average difference			-3.581

#### Table 11: Tests of Prediction Equality

*Notes:* The Normality test is based on the skewness and kurtosis of the distribution of the prediction differences. The test statistic is given in Doornik and Hansen (1994).

Source: Own calculations based on VLSS 1998, VHLSS 2002, VHLSS 2004 and the Population and Housing Census 1999.



Figure 1: Cross Plots of Provincial Poverty Incidence Predictions from the four Models

Source: Table 10.

Figure 1 shows cross plots of the four provincial level poverty predictions and Table 11 reports the three tests for equality of predictions described in Section 2. First, comparing the VLSS 1998 with the Minot and Baulch (2004) predictions it is clear from Panel A in Figure 1 that the association between the two predictions is very close and, accordingly, none of the three tests of prediction equality reject the null hypotheses of a zero median and a zero mean of the differences between the predictions.<sup>131</sup>

Second, comparing the VLSS 1998 and the VHLSS 2002 predictions the three test statistics also lead us to conclude that the predictions are equally good estimates of the underlying true poverty incidences. Panel B in Figure 1 shows a much larger dispersion of the predictions compared to the VLSS 1998 and Minot-Baulch pair, but the differences are not systematically different from zero when evaluated by the test statistics. The similarity of the VLSS 1998 and VHLSS 2002 predictions is confirmed by the comparison of the VHLSS 2002 and the Minot and Baulch predictions in Panel C in Figure 1 and in Table 11. The result of this comparison is also that we cannot reject that the predictions are from the same underlying distribution.<sup>132</sup>

Finally, when the predictions based on the VHLSS 2004 are compared with the three other sets we firmly reject equality and, again, all three tests lead to the same conclusion. A clear visual explanation for the rejection of equality can be found in Panel E in Figure 1 in which the predictions using the VHLSS 2004 are plotted against the predictions using the VHLSS 2002. The plot shows that the VHLSS 2004 poverty incidence predictions are almost all lower than the VHLSS 2002 predictions, and hence both the median and the mean of the differences is less than zero leading to a clear rejection of the null hypothesis of equality.

Table 12: Regression Based Tests of Prediction Equality

Dependent variable	Intercept	t-value	Slope	t-value	Joint sig.
P0(MB) - P0(VLSS 1998)	1.43	4.62	-0.03	-4.82	0.000
P0(VLSS 2002) - P0(VLSS 1998)	3.37	2.15	-0.06	-1.56	0.057
P0(VLSS 2004) - P0(VLSS 1998)	0.62	0.48	-0.08	-2.58	0.000
P0(Minot and Baulch) - P0(VLSS 2002)	1.78	0.97	-0.07	-1.66	0.076
P0(VHLSS 2002) - P0(VLSS 2004)	-1.82	-2.35	-0.04	-2.42	0.000
P0(Minot and Baulch) - P0(VLSS 2004)	3.32	2.25	-0.02	-0.59	0.000

Notes: The regression model is given in equation (18) in the main text.

Source: Own calculations based on Minot and Baulch (2004, Table 7.6), VLSS 1998, VHLSS 2002, VHLSS 2004 and the Population and Housing Census 1999.

<sup>&</sup>lt;sup>131</sup> The result of the *t*-test should be interpreted with caution though, as the hypothesis of normally distributed differences is firmly rejected.

<sup>&</sup>lt;sup>132</sup> In both of these comparisons we note again that the *t*-tests should not be overemphasized as the assumption of normally distributed prediction differences is rejected by the normality test.



Figure 2: Cross-Plots of Prediction Differences and Predictions

Source: Table 10.

Table 12 reports the results of the regression based comparisons of the predictions and Figure 2 gives the regression cross plots. Interestingly, the regression test of equality of the predictions based on VLSS 1998 and Minot and Baulch, respectively, rejects equality as we find a statistically significant systematic tendency of positive differences at low levels of poverty and negative differences at high

levels of poverty. This result is interesting because Minot and Baulch (2004, p. 241) note that their census-based poverty estimates tend to be less extreme than the survey based estimates when they are aggregated to a regional level. Our finding suggests that this result may be explained by the choice of estimator.

The other regression-based comparisons do not lead to changes in the conclusions reached above although equality of the VLSS 1998 and VHLSS 2002 predictions is only marginally accepted at the five percent level of significance. Moreover, the difference between the VHLSS 2004 and Minot and Baulch predictions appears to be a pure mean shift in the poverty incidences while the correlation between the predictions is 1. This is potentially interesting as it shows that a simple mean correction of the VHLSS 2004 predictions may be sufficient to obtain similar poverty incidence estimates.

Overall, the prediction plots clearly show that the point estimates of the provincial level poverty incidences differ, sometimes even quite substantially (*e.g.*, by 10 percentage points). But based on purely statistical measures the predictions using the VHLSS 2002 survey to estimate the parameters of a consumption model are as good as the predictions using the VLSS 1998 survey. In contrast, the predictions generated from the VHLSS 2004 are statistically different.

As a final assessment of the predictions we compare more aggregate poverty incidences derived from the provincial levels with purely survey-based estimates. Using the provincial population from the full Population and Housing Census as weights we estimate regional poverty incidences by averaging over the estimated provincial level poverty incidences. The resulting estimates of regional level poverty are given in Table 13.

	Census	Census predictions			Survey based estimates			
				Linear	Approx.		95%	
١		VHLSS	VHLSS	Interpolation	Confidence bounds			
Region	1998	2002	2004	1999	Lower	Upper		
Northern Mountainous R.	52.8	49.7	46.0	53.1	42.0	64.2		
Red River Delta	31.2	33.2	27.7	26.8	20.4	33.3		
North Central Coast	47.6	56.9	51.4	47.0	36.6	57.4		
South Central Coast	39.4	40.7	36.5	32.2	20.9	43.4		
Central Highlands	45.1	53.2	45.8	52.5	33.3	71.7		
South East	6.9	10.6	8.8	7.9	5.0	10.8		
Mekong River Delta	35.1	32.0	31.7	32.7	26.7	38.7		
Vietnam	36.3	37.7	34.2	34.8	31.6	36.1		
Prediction Statistics								
Mean prediction error <sup>a</sup>	-0.8	-3.4	0.6					
RMSE <sup>b</sup>	4.4	5.7	4.4					
MAE <sup>c</sup>	3.3	4.6	3.6					

Table 13: Estimated Regional Poverty Incidence (Percent)

*Notes:* The census predictions are aggregated from the provincial level predictions using the provincial populations as weights. The population statistics are from the full census. The survey based estimates are estimated by linear interpolation between the regional poverty incidences in the VLSS 1998 and the VHLSS 2002 using the comparable consumption measure as poverty indicator. <sup>a</sup> Mean prediction error is the average of the seven regional prediction errors using the survey based poverty incidences. <sup>b</sup> RMSE is the root mean squared prediction error for the seven regions. <sup>c</sup> MAE is the mean absolute error for the seven regions.

Source: Own calculations based on VLSS 1998, VHLSS 2002, VHLSS 2004 and the Population and Housing Census 1999.

Next, the problem is to find a suitable survey based estimate. The census based estimates in Table 13 are representing the regional poverty as of April 1, 1999. Hence, the VLSS 1998 cannot be used directly for comparison because this is an estimate of the poverty as of January 1, 1998. To get survey based regional poverty incidences for 1999 we may assume the annual change in the regional poverty incidence has been constant from 1988 to 2002. Given this assumption the poverty incidences for 1999 can be estimated by linear interpolation using the regional level poverty incidences from VLSS 1998 and VHLSS 2002. The results of the linear interpolations are given on the right hand side of Table 13.<sup>133</sup> In addition to the interpolated point estimates the approximate 95 percent confidence bounds are also reported to show the degree of uncertainty in the survey based estimates.

The three poverty map predictions all aggregate to regional level estimates that are within the 95 percent confidence bounds of the survey based estimates. But at the most aggregate level (the poverty incidence for Vietnam) the predictions based on VLSS 1998 and VHLSS 2002 are above the upper confidence bound, *i.e.*, they both predict too high poverty in 1999 compared to the linear interpolation. Looking at the prediction errors we find that all three census based predictions underestimate the poverty in the Northern Mountainous Region, while they overestimate the poverty in the Red River Delta, the North Central Coast and the South Central Coast. In the three remaining regions (the Central Highlands, the South East and the Mekong River Delta) there is variation in the sign of the prediction errors.

Three summary statistics of the prediction errors for the seven regions are given in the bottom part of Table 13. The mean prediction error shows that the VLSS 1998 and the VHLSS 2004 are both practically unbiased as they are within one percentage point from the survey based estimates. The mean prediction error is somewhat higher for the VHLSS 2002 prediction but given the uncertainty in the estimates it is not excessive.

The most interesting result in Table 13 is that the predictions based on the VHLSS 2004 are as good as the predictions based on the VLSS 1998 when they are evaluated using standard measures such as the root mean squared prediction error and the mean absolute prediction error. Even though these summary statistics can only be indicative—being based on seven observations—they illustrate the problems in evaluating the poverty mapping results. The analysis above shows that the two census predictions using the VLSS 1998 and VHLSS 2004 for the consumption model are statistically different. But based on the results in Table 13 we cannot conclude that the VLSS 1998 predictions are better than the VHLSS 2004.

Overall, the finding that the three census based predictions result in somewhat different poverty estimates at the regional level is not surprising and the simple evaluation in Table 13 does not lead to a firm conclusion as to whether any of the three prediction models should be rejected. Finally, regarding the results from the VHLSS 2002 predictions we note that the GSO in Vietnam is currently discussing if the consumption expenditure in the survey should be adjusted upwards by 4 percent to better match the national accounts data. Such an adjustment would lead to lower estimates of the provincial and regional poverty incidence in the poverty map model and that would probably improve the prediction statistics in Table 13. But given that the precise adjustment is not known at the time of writing we have not included the adjustment in our analysis.

## 6. Conclusion

During the last decade poverty mappings have become an increasingly popular tool in investigations

<sup>&</sup>lt;sup>133</sup> We have also experimented with non-linear interpolation in which we assume the elasticity of the regional poverty incidences with respect to national GDP growth is constant over time. The results were close to the ones reported in Table 13.

of social, environmental and economic problems in many developing countries and currently about 25 developing countries have produced detailed poverty maps to assist local and central policy makers.

Vietnam has poverty maps with estimates of the poverty incidence at the commune level (Minot, Baulch and Epprecht, 2003) and at the provincial level (Minot and Baulch, 2004). These maps are estimated by combining the Vietnamese living standard measurement survey from 1998 and the 1999 Population and Housing Census. Because of the rapid socio-economic development in Vietnam there is a growing interest in updating these poverty maps to more recent years. However, even though there are new living standard measurement surveys (VHLSS 2002 and VHLSS 2004) a new census will not be available until the end of the decade. Therefore, to update the poverty maps other solutions than waiting for a new census must be investigated. One possibility is to update the poverty maps by using a panel component of two surveys and this option is available in Vietnam as the VHLSS 2002 and VHLSS 2004 have a panel component of about 4,000 households. However, a first step in such a panel based update of the poverty map is to assess if the VHLSS 2002 in combination with the Population and Housing Census from 1999 provides good estimates of the local area poverty incidences.

In the present study we have made a first pass by comparing provincial level poverty incidences, which are based on consumption models using the three surveys VLSS 1998, VHLSS 2002 and VHLSS 2004. We first compare the population estimates of the regressors in the consumption function across the three surveys and the census. Surprisingly, the result is that, in general, none of the three surveys match the census based population means. Hence, it is not clear that any of the three surveys represent the same underlying populations as the census. While this is somewhat unexpected considering that the time spread between the VLSS 1998 and the census is only one year, it need not invalidate the poverty mapping results. But, the result implies that there are no strong statistical reasons for preferring the VLSS 1998 in the poverty mapping instead of the VHLSS 2002. In contrast, one may argue in favor of the VHLSS 2002 because of the much larger sample size that allows for a more detailed consumption model.

The second test of the poverty mapping models is a test of parameter constancy across consumption models, using data from each of the three household surveys. We find that the consumption model parameters are statistically different when they are estimated using different surveys. This means that the systematic components captured in the consumption model are changing over time. This may be interpreted as a clear rejection of a constant underlying super-population.

Because the poverty incidence predictions are non-linear functions of the consumption function parameters, non-constancy of the latter parameters does not imply that the poverty predictions will be statistically different. Therefore, we use the each of three surveys in combination with the census to estimate provincial level poverty incidences. Subsequently, we test if the predicted poverty incidences are statistically different. It turns out that the predictions based on the combination of the VLSS 1998 and the census, are not statistically different from the predictions generated from the combination of the VHLSS 2002 and the census. In contrast we find the last combination, the VHLSS 2004 and the census to result in statistically different predictions. Based on the statistical analysis we conclude that the combination of the VHLSS 2002 and the census generates poverty incidence predictions that appear to be valid estimates of the poverty situation in 1999.

The conclusion opens up for further poverty mapping analyses using the VHLSS 2002. First of all, it would be interesting if the same conclusion is reached at the commune level, *i.e*, by combining the survey with the 33 percent sample of the Population and Housing Census 1999. Second, it allows for more flexibility in the consumption models because the sample in the VHLSS 2002 is very large. In particular, it seems worthwhile to analyze if allowing for parameter variation across ethnic majority/minority groups can improve the predictions. Furthermore, stratum level estimations of the consumption function parameters can be estimated with greater precision, and this may also result in

improved predictions. Finally, the poverty map can possibly be updated by combining the data in the VHLSS 2002 and the VHLSS 2004 as discussed above.

## References

- Demombynes, G., Elbers, Lanjouw, J. O., Lanjouw, P., Mistiaen, J. A. and Özler, B. 2003. "Producing an Improved Geographic Profile of Poverty," in R. van der Hoeven and A. Shorrocks (eds.) *Growth, Inequality and Poverty.* Oxford: Oxford University Press.
- Demombynes, G., Elbers, Lanjouw, J. O., Lanjouw, P., Mistiaen, J. A. and Özler, B. 2002. "Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries". WIDER Discussion Paper #39: United Nations University.
- Doornik, J. and Hansen, H., 1994. "An omnibus test for univariate and multivariate normality". Working Paper, Nuffield College, Oxford.
- Elbers, C., Lanjouw, J. and Lanjouw, P., 2000. "Welfare in villages and towns: micro-level estimation of poverty and inequality." Discussion Paper TI 2000-029/2. Tinbergen Institute.
- Elbers, C., Lanjouw, J. and Lanjouw, P., 2002. "Micro-level estimation of welfare." Policy Research Department Working Paper No. WPS2911. The World Bank.
- Elbers, C., Lanjouw, J. and Lanjouw, P., 2003. "Micro-level estimation of poverty and inquality." *Econometrica* 71(1): 355-364.
- Ghosh, M. and J.N.K Rao. 1994. "Small area estimation: an appraisal". Statistical Science, 9(1), 55-93.
- Glewwe, P. 2004. "An overview of economic growth and household welfare in Vietnam in the 1990s" Chapter 1 in P. Glewwe, N. Agrawal and D. Dollar, *Economic Growth, Poverty, and Household Welfare in Vietnam.* Regional and Sectoral Studies, Washington DC: World Bank.
- Henninger, N. and M. Snel. 2002. Where are the poor? Experiences with the Development and Use of Poverty Maps. Washington, DC.: World Resources Institute.
- Hentschel, J., Lanjouw, J., Lanjouw, P. and Poggi, J., 2000. "Combining census and survey data to trace the spatial dimensions of poverty: a case study of Ecuador", *World Bank Economic Review*, 14(1): 147-65.
- Hoogeveen, J. G., T. Emwanu, and P. O. Okwi. 2003. "Updating small area welfare indicators in the absence of a new census". Working paper, November 13, 2003. Processed.
- Inter-Ministerial Poverty Mapping Task Force. 2003. Poverty Mapping and Market Access in Vietnam. Hanoi.
- Lanjouw, P. 2003. "Estimating geographically disaggregated welfare levels and changes". In: F. Bourguignon and L. A. Pereira da Silva (eds.), *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. A copublication of the World Bank and Oxford University Press.
- Minot, N. and B. Baulch. 2004. "The spatial distribution of poverty in Vietnam and the potential for targeting." In P. Glewwe, N. Agrawal and D. Dollar, *Economic Growth, Poverty, and Household Welfare in Vietnam*. Regional and Sectoral Studies, Washington DC: World Bank.
- Minot, N., B. Baulch and M Epprecht. 2003. Poverty and inequality in Vietnam: Spatial patterns and geographic determinants. Hanoi, Vietnam.
- Rao, J.N.K. 1999. "Some recent advances in model-based small area estimation". Survey Methodology, 25(2), 175-186.
- Rao, J.N.K. 2003. Small Area Estimation, New York: Wiley.

World Bank. 1999. Vietnam Development Report 2000: Attacking Poverty. Hanoi, Vietnam

- World Bank. 2001. "Vietnam living standards survey (VLSS) 1997-98: basic information". Poverty and Human Resources Division, The World Bank. Processed.
- World Bank. 2003. Vietnam Development Report 2004: Poverty. Hanoi, Vietnam.

## Appendix

Difference between	1998 and 2002		1998 and 2004		2002 and 2004	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Household size	0.037	2.43	0.045	2.53	0.007	0.55
Household size squared	-0.001	-1.15	-0.003	-1.77	-0.001	-1.01
Proportion of elderly people (>60)	0.004	0.10	0.015	0.34	0.011	0.33
Proportion of children (<15)	-0.143	-3.54	-0.186	-4.16	-0.043	-1.33
Proportion of females	0.054	1.44	0.112	2.56	0.058	1.76
Ethnic minority head	-0.093	-2.10	-0.148	-3.20	-0.054	-2.08
Primary school	-0.002	-0.10	0.007	0.33	0.009	0.54
Lower secondary school	-0.015	-0.68	0.045	1.76	0.060	3.04
Upper secondary school	0.019	0.62	0.036	1.02	0.017	0.62
Technical or vocation training	0.036	0.93	0.066	1.44	0.031	0.77
College diploma/University degree	-0.036	-0.58	-0.025	-0.38	0.011	0.21
No spouse	0.039	1.80	0.038	1.47	-0.001	-0.03
Primary school	0.017	0.83	0.011	0.48	-0.006	-0.32
Lower secondary school	0.021	0.82	0.006	0.21	-0.015	-0.71
Upper secondary school	0.089	2.06	0.034	0.72	-0.055	-1.64
Technical or vocation training	0.140	3.01	0.104	1.91	-0.036	-0.78
College diploma/University degree	0.164	2.37	0.114	1.57	-0.050	-0.85
Leaders in all fields	-0.035	-0.73	-0.083	-1.44	-0.048	-0.98
Professionals in all fields	-0.020	-0.40	-0.023	-0.41	-0.003	-0.06
Clerk or service workers	-0.041	-0.89	-0.066	-1.29	-0.025	-0.56
Skilled workers in agriculture	0.001	0.03	-0.034	-1.03	-0.035	-1.44
Skilled workers	0.009	0.23	-0.047	-1.05	-0.056	-1.80
Unskilled workers and army force	0.130	3.53	0.059	1.45	-0.071	-2.53
Permanent house	0.513	2.25	0.155	0.61	-0.358	-2.38
Semi-permanent house	-0.097	-1.00	-0.140	-1.24	-0.043	-0.52
Permanent house x log(area)	-0.141	-2.54	-0.058	-0.93	0.083	2.32
Semi-permanent house x log(area)	0.019	0.75	0.033	1.14	0.014	0.69
Electricity	-0.012	-0.37	-0.019	-0.54	-0.008	-0.30
Private or public tap water	0.050	0.79	0.024	0.38	-0.026	-0.73
Rain water and wells	-0.067	-2.34	-0.087	-2.87	-0.020	-1.03
Flush toilet	-0.019	-0.37	-0.063	-1.20	-0.044	-1.58
Latrine/other toilet	-0.051	-2.16	-0.048	-1.83	0.003	0.17
TV ownership	-0.011	-0.68	0.012	0.62	0.023	1.56
Radio ownership	-0.010	-0.62	-0.006	-0.31	0.004	0.32
Red River Delta	-0.090	-1.61	-0.082	-1.44	0.008	0.31
North Central Coast	-0.170	-2.68	-0.173	-2.70	-0.004	-0.15
South Central Coast	-0.115	-1.79	-0.150	-2.28	-0.035	-1.14
Central Highlands	-0.185	-2.29	-0.143	-1.70	0.043	1.11
South East	-0.267	-4.41	-0.264	-4.24	0.004	0.10
Mekong River Delta	-0.077	-1.27	-0.151	-2.44	-0.074	-2.69
Constant	0.064	0.75	0.152	1.61	0.088	1.46
Chow-test (p-value)	0.000		0.000		0.000	
Chow-test of slopes (p-value)	0.000		0.000		0.005	
Chow-test of regions and intercept (p-value)	0.000		0.002		0.032	

Notes: The dependent variable is log real per capita expenditure in January 1998 prices. The regressions are based on weighted least squares using the survey specific sampling weights. The *t*-statistics are cluster robust.

Source: VLSS 1998, VHLSS 2002 and VHLSS 2004.

Difference between	1998 and 2002		1998 and 2004		2002 and	2004
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Household size	0.040	1.82	0.020	0.71	-0.020	-0.72
Household size squared	-0.002	-1.10	0.001	0.52	0.003	1.26
Proportion of elderly people (>60)	0.009	0.11	0.087	1.01	0.079	1.09
Proportion of children (<15)	-0.141	-1.84	-0.066	-0.78	0.075	1.15
Proportion of females	-0.097	-1.08	-0.185	-1.96	-0.089	-1.37
Ethnic minority head	-0.192	-1.78	-0.212	-1.89	-0.020	-0.25
Primary school	-0.038	-0.84	0.006	0.13	0.044	1.03
Lower secondary school	0.069	1.72	0.138	3.21	0.069	1.60
Upper secondary school	0.020	0.37	0.077	1.28	0.057	1.08
Technical or vocation training	0.030	0.54	0.075	1.18	0.045	0.82
College diploma/University degree	0.100	1.57	0.152	2.20	0.052	0.83
No spouse	0.055	1.14	0.051	0.99	-0.004	-0.08
Primary school	-0.043	-0.99	-0.097	-2.04	-0.054	-1.20
Lower secondary school	-0.045	-0.97	-0.094	-1.96	-0.050	-1.13
Upper secondary school	-0.061	-0.77	-0.142	-1.70	-0.081	-1.40
Technical or vocation training	-0.015	-0.29	-0.002	-0.03	0.013	0.24
College diploma/University degree	0.077	1.07	0.022	0.29	-0.055	-0.86
Leaders in all fields	-0.153	-1.53	-0.070	-0.71	0.083	0.92
Professionals in all fields	0.002	0.04	-0.032	-0.50	-0.034	-0.66
Clerk or service workers	0.015	0.30	-0.048	-0.88	-0.063	-1.27
Skilled workers in agriculture	-0.055	-1.03	-0.017	-0.28	0.038	0.91
Skilled workers	-0.020	-0.42	-0.030	-0.61	-0.011	-0.26
Unskilled workers and army force	0.157	2.69	0.147	2.37	-0.010	-0.26
Permanent house	0.132	0.70	-0.066	-0.32	-0.198	-1.13
Semi-permanent house	0.036	0.26	0.035	0.22	-0.001	-0.01
Permanent house x log(area)	-0.024	-0.54	-0.001	-0.03	0.023	0.57
Semi-permanent house x log(area)	-0.011	-0.31	-0.032	-0.78	-0.020	-0.58
Electricity	-0.045	-0.46	0.137	1.07	0.182	1.39
Private or public tap water	-0.052	-1.00	-0.104	-1.85	-0.053	-1.01
Rain water and wells	-0.056	-0.93	-0.074	-1.14	-0.018	-0.35
Flush toilet	0.054	0.73	0.083	1.01	0.030	0.47
Latrine/other toilet	-0.072	-1.15	-0.120	-1.64	-0.047	-0.80
TV ownership	0.025	0.54	0.047	0.92	0.022	0.49
Radio ownership	-0.031	-0.98	-0.033	-0.95	-0.002	-0.07
Red River Delta	0.001	0.01	0.044	0.43	0.044	0.79
North Central Coast	-0.178	-1.93	-0.164	-1.73	0.014	0.27
South Central Coast	-0.178	-2.08	-0.105	-1.19	0.073	1.33
Central Highlands					0.070	0.96
South East	-0.020	-0.24	0.031	0.36	0.051	0.94
Mekong River Delta	-0.136	-1.42	-0.096	-0.96	0.040	0.72
Constant	0.176	1.08	0.108	0.54	-0.068	-0.35
Chow-test (p-value)	0.000		0.000		0.427	
Chow-test of slopes (p-value)	0.000		0.000		0.487	
Chow-test of regions and intercept (p-value)	0.074		0.144		0.934	

 Table A 2: Chow-Tests of Parameter Constancy in the Consumption Model Urban Area

Notes: The dependent variable is log real per capita expenditure in January 1998 prices. The regressions are based on weighted least squares using the survey specific sampling weights. The *t*-statistics are cluster robust.

Source: VLSS 1998, VHLSS 2002 and VHLSS 2004.

Difference between	1998 and 2002		1998 and	2004	2002 and 2004	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
(xb) <sup>2</sup>	0.0089	1.50	0.0123	1.36	0.0034	0.62
(xb) <sup>2</sup> x ethnic, rural	0.0036	2.32	0.0051	2.15	0.0015	0.98
(xb) <sup>2</sup> x Household size, rural	-0.0011	-1.15	-0.0018	-1.25	-0.0007	-0.79
(xb) <sup>2</sup> x Household size sq., rural	0.0001	1.24	0.0002	1.36	0.0001	0.85
(xb) <sup>2</sup> x Household size, urban	-0.0023	-1.85	-0.0027	-1.32	-0.0004	-0.28
(xb) <sup>2</sup> x Household size sq., urban	0.0002	1.68	0.0000	0.26	-0.0001	-0.97
(xb) <sup>2</sup> x Urban	0.2417	1.43	0.1963	0.72	-0.0453	-0.31
(xb) x Urban	-4.3022	-1.53	-3.6868	-0.81	0.6155	0.24
Urban	19.2891	1.65	17.2790	0.91	-2.0101	-0.19
Constant	-0.9906	-2.86	-0.5216	-0.96	0.4690	1.40
Chow test (p-value)	0.000		0.000		0.000	

Table A 3: Chow-Tests of Parameter C	Constancy in the Variance Model
--------------------------------------	---------------------------------

*Note:* The dependent variable is the logistic transformation of the squared residuals from the consumption regressions as shown in equation (X). The regressions are based on weighted least squares using the survey specific sampling weights. (xb) are the fitted values from the consumption regressions in Table and Table .

Source: VLSS 1998, VHLSS 2002 and VHLSS 2004.

## Annex

# Vietnam Household Living Standards Survey (VHLSS) 2002 and 2004

Basic Information Phung Duc Tung and Nguyen Phong

## 1. Introduction

During the 1990s two important household surveys were developed in Vietnam to look at living standards: the Multi-purpose Household Survey (MPHS) and the Vietnam Living Standards Survey (VLSS). The MPHS is a key component of the overall statistical system and survey program as information collected is used for multiple purposes including CPI calculations, national account statistics and poverty estimates among others.

The MPHS survey has been implemented every 1 to 2 years since 1994 and has a sample size from 25,000 to 47,000 households. The focus of the survey has been household income and expenditures; however other important aspects of living standards such as education, health or employment have been included in various rounds. The large sample size, limited budget, large number of personnel involved, short time for implementation, unclear medium and long-term plans, significant changes in methodology and content over time and lack of standardization in methodology have led to some inconsistencies in data collection over time. Furthermore, inability to make clear analysis plans has possibly had negative impacts on data quality. In addition, the MPHS has not been disseminated in a timely manner nor analyzed in depth.

The two VLSS were implemented in 1992-93 and 1997-98 with funding from the UNDP and Sida and with technical assistance from the World Bank. They provided important opportunities for the General Statistics Office (GSO) to learn a new survey methodology of integrating many topics into one survey as well as exposure to up-to-date methods of questionnaire design and quality control. It also provided a high quality data set covering a great variety of topics that has been widely used in social policy research. However, with the given budget and a long, detailed questionnaire only a relatively small sample (4,800 households in 1992-93 and 6,000 households in 1997-98) could be selected, which was not large enough to provide information at a level that satisfied the government or provinces. In addition, the frequency of implementation every five years was too low to satisfy the need for up-to-date information to monitor social issues.

The implementation of these two different surveys with similar purposes has led to duplication and waste. It is therefore a high priority of the GSO to integrate the MPHS and the VLSS surveys and create a medium term plan for implementing a new VHLSS through 2010. With the technical assistant from UNDP and World Bank, GSO has developed a strategy for the VHLSS over the next 10 years (2000-2010) that transformed the VHLSS into a biennial (once every two years) core and rotating module household survey. The VHLSS strategy is approved and it will be implemented as follows:

Year	Survey content
2002	Expenditure and income (core) + basic information of other sections
2004	Core + Land and Non farm activities (rotating module)
2006	Core + Health and Education (rotating module)
2008	Core + Land and Non-farm activities
2010	Core + Health and Education

This document provides background information on the Vietnam Household Living Standards Survey 2002 and 2004, and the data-sets resulting from this nation-wide household survey. Information is provided on the survey design and implementation, the content of the different questionnaires, data processing activities and data dissemination policy for this survey.

## 2. The Survey Questionnaires

## 2.1. The VHLSS 2002 Questionnaires

The VHLSS 2002 used 3 questionnaires: a short household questionnaire (excluding most of the consumption expenditure information), a long household questionnaire (including detailed consumption expenditure information), and a commune questionnaire.

### 2.1.1. The Household Questionnaire

The short household questionnaire contained 9 sections, each of which covered a separate aspect of household activity.

cover page (survey information)

- 1. Household roster
- 2. Education
- 3. Employment
- 4. Health
- 5. Income and household production
- 6. Expenditure (the short questionnaire collected only some main food expenditure)
- 7. Durable goods and assets
- 8. Housing
- 9. Participation in poverty reduction programs

The individual designated by the household members as the household head provided responses. For some questions (on education and health expenditure, income from wage, agro-pastoral activities, non-farm self-employment, food expenditures, non-food expenditures) a member identified as the most knowledgeable in each of those sections provided responses.

Unlike the VLSS, the household questionnaire of VHLSS 2002 was completed in one interview. The survey was designed so that more sensitive issues such as credit, savings and assets were discussed near the end. The content of each module is briefly described below.

## SECTION 0: COVER PAGE

The cover page collected information about the household identification codes (province, district, commune, enumeration area and household codes), the religion, ethnic group of the household,

language used by the respondent, and codes of interviewer and team leader as well as the date of interview.

## SECTION 1: HOUSEHOLD ROSTER

The household roster lists the sex, relation to head of household, birth dates, age and marital status (for people aged 13 and over) of all people who are living in the household since at least 6 months ago. The household head is listed first and receives the personal id code 1. Household members were generally defined as "all people who normally live and eat their meals together in this house and have done so for 6 or more months out of the past year". However, there are 7 exceptions:

1. The head of household is always considered as a household member even if he or she does not live and eat in the household more than 6 months.

2. Infants less than 6 months old are still considered as household members.

3. The people who are going to live in the household for a long time such as daughter-in-law, son-in-law, relatives who came back from outside (retired, soldier, etc.) are considered as household members even if they have been living less than 6 months in the household.

4. Students living outside the household but still supported by their family are considered as household members.

5. Guests or relatives living with the household 6 or more months and eat their meals together are considered as household members.

6. Hired workers, servants, or lodgers or guests are not the household members if they are members of households elsewhere.

7. Individuals who died during the past 12 months or moved out of the household and do not intend to come back are not considered as household members.

A transforming table between lunar and solar year is provided in the questionnaire to help respondents recall the year they were born.

#### SECTION 2: EDUCATION

Section 2 first asks which grade an individual has completed in the school. If the grade reported is less than 5, the section then asks whether she or he can read and write, the highest obtainable diploma and if they are currently in school or have completed their schooling. Depending on the answer, the interview continues to ask details about school expenditures including tuition, contribution (construction fund, parent's association), uniforms, textbooks, other education equipments, extra learning (including foreign language and computer), other expenses and the total school expenditure. Individuals were asked whether or not they were allowed exemptions from or reductions in fees, and for what reason as well as the percentage of exemption for tuition and contributions. They are also asked about the value of scholarships. The last question in this section is to ask individuals about the expenditure in the past 12 months on other courses outside of school such as study of foreign languages or computers.

#### **SECTION 3: EMPLOYMENT**

All individuals whose ages are 10 and older are asked to respond to the economic activity questions in Section 3, beginning with questions on the nature of their works in the last 12 months. First question asks whether an individual has been worked as a wage earner or self-employed in agriculture or self-employed in non-farm activities. If they have worked in the past 12 months, the next question is on working hours in the past 7 days and the main job in the past 12 months. Information about occupation, industry of employment, type of employer, number of working months in the past 12 months, average number of working days per months in these months, and average number of

working hours per day are collected. Then the same information about total number of working months, days per month and hours per days of total other earned jobs are collected. Finally, hours spent doing household chores per day and numbers of days of community work are collected for each household member aged 10 and older. Occupation and industry of employment codes are printed directly in the household questionnaire.

## SECTION 4: HEALTH

The section begins by asking respondents whether or not any individual in their family went to any health provider in the past 12 months (including a health provider coming to their home, health examination, vaccination with payment, pregnant examination, child delivery), and if yes, the next questions asks about the name of the health providers, inpatient or outpatient and the total cost of each time for each inpatient or outpatient treatment. The rest of the questions are asked for all household members about other health expenditures on medicine and health equipment, contribution for health fund for local community and expenditure on health insurance in the past 12 months.

## SECTION 5: INCOME AND HOUSEHOLD PRODUCTION

5A	Income from wage
5B	Farm activities
5B1	Agriculture, forestry and water- surface lands
5B2	Crops
5B2.1	Food crops
5B2.2	Annual and perennial industrial crops
5B2.2	Fruit crops
5B2.4	Crop by-products
5B2.5	Crop expenditures
5B3	Livestock
5B3.1	Turnover from livestock
5B3.2	Livestock expenditure
5B4	Agricultural services
5B4.1	Output from agricultural services
5B4.2	Expenditure for agricultural service activities
5B5	Forestry activities
5B5.1	Output from forestry activities
5B5.2	Expenditure on forestry activities
5B6	Fishery activities
5B6.1	Output from fishery activities
5B6.2	Expenditure on fishery activities
5C	Non-farm activities
5C1	Output from non- farm activities
5C2	Expenditure on non- farm activities
5D	Other income sources
5D1	Other income
5D2	Non-income sources

In this section, most questions refer to the past twelve months. This section is the longest one in the household questionnaire, with many subsections that contain information on different aspects of income sources from wage, agricultural production and related livestock and fishery and non-farm activities.

Section 5A collects information of the household member who has been working as employee in the past 12 months, starting with the questions on occupation, industry of employment and type of employer, total salary received and other incomes (received during Tet holiday, social subsidy for working accident, during pregnancy time, the accommodation in business trips and others) for the main job during the past 12 months of the main job. The last two questions in this section ask whether the individual has been working as an employee in a second or subsequent jobs, and if yes, what is the total amount that she or he received during the past 12 months.

Section 5B1 collects information on household's control over different plots of land of different tenures. These include their own land, auction land, rented land, and land rented out by the household. Annual crop and perennial crop land, forestry land, water surface and unused land are included. In each case data are obtained on total area, area for long use purpose, area with land use certificate (LUC) and who has the name on the LUC.

Sections 5B2.1 to 5B2.3 contain detailed information on output of all crops grown by the household. The information is obtained separately for each crop and includes (in most cases) information on the total area planted, quantity harvested, amount,,value sold and the main buyer, and total value of output in the past 12 months.

Section 5B2.4 contains information about total value of crop by-products during the past 12 months, including straw, thatch, sweet potato leaves and sterms, sterms of cassava plant, etc.

Section 5B2.5 contains detailed information on the use of production inputs including seeds, manure, chemical fertilizer, insecticides and herbicides, small equipment, fuel including electricity, gasoline, repair and maintenance, depreciation of equipment and assets, expenditure on rented land and equipments, labor, irrigation fee, interest payment on loans, agricultural tax, and other expenditures (fee, post, insurance, etc.) for the 4 main groups of all crops cultivated by the household. This information is crop-specific for each main group of crops, so it can be linked with the output information in the earlier sections.

Section 5B3.1 collects information on the output of livestock, poultry and other animals that are sold, consumed by a household, or used for other purposes, the total amount, and the value of output in the past 12 months.

Section 5B3.2 collects information on the cost of livestock, poultry and other animals specified by animal feed, equipment, fuel, repair and maintenance, depreciation of assets, renting land, labor, medicine, interest payment, tax and other expenditures.

Section 5B4.1 and 5B4.2 collect information on the revenue and expenditure of agricultural service activities of the household. These activities include plowing/land preparation service, irrigation, primary crop processing service and other services.

Section 5B5.1 and 5B5.2 collect information about the output and cost of forestry activity of the household in the past 12 months for 10 main types of tree including mu oil tree, cinnamon tree, anise tree, pine tree, varnish tree, tree for wood, bamboo, fan falm tree, water coconut falm, other forestry tree and wood. The cost items in section 5B5.2 are similar to the section 5B2.1.

Section 5B6.1 and section 5B6.2 collect information about the output and cost of aquaculture production in the past 12 months including the quantity and value the household consumed and sold, the cost of raising, catching, and service for breed, foods, non-durable items, energy fuel, small repair, etc.

Section 5C1 collects information about non-farm activities of the household in the past 12 months, including the type of activity, industry code, the main person who operated, turnover, values the household consumed in the past 12 months, and the cost for each activity including material, electricity, small repair and maintenance, depreciation of assets, labor cost, interest payment, tax, etc.

Section 5D1 collects information about income from other sources in the past 12 months. The other household income includes remittance and value of in- kind received from both domestic and

overseas individuals and organizations, pension, sickness and one-time job loss allowance, social insurance allowance, other income from social insurance, interest of saving, dividend, income from renting out workshop, machinery, assets, income from wining lottery, from charity, and support from other organizations.

Section 5D2 collects information on the sale of machines, equipment, workshop, house, gold, and withdrawal from saving, stock, etc.

### **SECTION 6: EXPENDITURE**

Section 6 in the short questionnaire collects only 13 key food items the household consumed in the past 12 months. Those food items include pork, beef, chicken, all kinds of fish, shrimp, fruit, cakes, milk, beer, coffee, bottled water and food eaten outside.

Section 6 in the long questionnaire is divided into two sections. Section 6 part A collects information about expenditure of the household on food and drink during holidays (section 6 part A1) and daily (section 6 part A2).

Section 6 part A1 collects information about quantity and value of 24 main food items that Vietnamese households usually consume during the holidays, especially in the Tet holiday (new lunar year). Those food items are mainly typical meats such as pork, beef, chicken; special and glutinous rice; typical drink such as coffee, tea, beer, wine and alcohol, fruit, and food bought outside or self made or received by the households during the holidays in the past 12 months.

Section 6 part A2 collects information about the household expenditure on daily food and drink of 58 food items, excluding the amount the household consumed during holidays. It collects detailed information on market purchases and consumption from home production. Information is obtained on the number of months (in the past 12 months) each food item was purchased, the number of times purchases were made during those months, the quantity purchased each time, and the value per purchase. These four pieces of information can be combined to obtain the total expenditure on food in the 12 months before the date of the interview. Besides market purchases (including barter), information is also collected on consumption from home production. Again data are obtained on the number of months each item was consumed, but unlike market purchases, the information of the quantity and value of consumption is obtained by asking a single question on the total amount for the last 12 months.

Section 6 part B collects information on non-food household expenditures.

In section 6 part B1 respondents were asked to recall the number of months purchased, the number of time purchases were made during these months, the total amount spent in the past 12 months as well as the value of self making on daily expenses such as lottery tickets, cigarettes, soap, personal care products, cooking fuel, matches and candles, gasoline, etc.

Section 6 part B2 collects information about annual expenditure for 32 items such as fabric, ready made clothing, mosquito net, face towel, scarves, rush mats, blankets, pillows, tailoring or laundry service, shoes, nylon sheeting, light buds, electric wire, etc. Data on the value purchased and self-making are also obtained for the past 12 months.

Section 6 part B3 collects information about the amount of money the household spent during the past 12 months for contributions to various funds, public labor contribution, all kinds of taxes (excluding production taxes), wedding, funeral, on special occasions of household members, parties (celebrate birthday), give, donate, support (cash and kind) and others.

Section 6 part B4 collects information on other household expenses that are not considered as household expenditure in the past 12 months, including lending money, debt payment, return advance (including payment of interest), contribution to revolving credit group, buying share, certificate and stock, purchase gold, silver, precious stone, foreign currency for saving purpose, saving account, life and security insurance, outstanding investment, and other expenditures.

## SECTION 7: FIXED ASSETS AND DURABLE GOODS

This section collects information about 59 kinds of fixed assets and durable goods of the households, including perennial crops garden, aquaculture production area, other production land area, buffalo, cow, horse for production and breeding, feed grinding machine, rice milling machine, car, trailer, motorbike, wagon, boat with engine, computer, television, telephone, air-conditioner, etc. For each asset or good questions are asked about the month and year the household received the asset/good, the value at the time of purchase or reception, the value at the time of interview, the using purpose, the (percentage) share using for production purpose and the share using for consumption purpose.

### SECTION 8: HOUSING

This section collects information about the living conditions of the household, including the number of houses or flats that the household is living in, the total living area, type of the main dwelling (quality of the house), the time of living in those houses or flats, the place the household had been living before the current place, the owner status of the current living place, amount of money paid during the past 12 months for renting of the houses or flats, the value of the house or flat at the interview time. In addition, this section also collects information about the other houses or flats the household own and whether the household received money from renting it out, the amount received from renting in the past 12 months and the value of that house or flat at the interview time. The household is also asked to provide information about expenditure on construction, repair, renovation and improvement of the house during the past 12 months, the main drinking source and type of toilet that the household is currently using, the amount paid for using the drinking water and the main source of lighting and the method the household uses to dispose the garbage in the past 12 months. Furthermore, this section collects information about the access of the household to Vietnamese television, radio channels and the popular newspapers in Vietnam as well as access to internet in the past 12 months.

## SECTION 9: PARTICIPATION IN POVERTY REDUCTION PROGRAMS

This section collects information with the purpose of examining the coverage of National Target Program to the poor households and their accession to the credit and health programs. In the first three questions, the household is asked to define whether the household belongs to a family of invalids, sick war veterans, martyr, Vietnamese heroic mother or lonely elderly, disabled or poor household and then to examine whether the household received the Poor Household Certificate or not. In addition, if the household received the Poor Household Certificate, it is asked to define whether they have got exemption from health consultation and treatment or received assistance from the State mass organizations, associations for repairing or constructing dwelling, or they have got exemption from agriculture land use tax in year 2001. The information about the loans, value of the loans, the interest rate from different banks and social funds, private borrowers are collected in the last 4 questions of this section.

#### 2.1.2 The Commune/Ward Questionnaire

The commune questionnaire is administered by the team leader and completed with the help of village chiefs, teachers, government officials and health care workers. The questionnaire is administered in both rural and urban areas but some sections are only collected in the rural area such as non-farm employment opportunities and infrastructure and transportation. The commune questionnaire contains 9 sections including:

- 0. survey information
- 1. main charateristics of the commune/ward
- 2. general economic conditions and aid programs

- 3. non-farm employment opportunities
- 4. agriculture
- 5. physical infrastructure and transportation
- 6. education
- 7. health
- 8. public disorder and other social affairs

Section 0 contains basic background information on respondents including position in the commune, tenure in the current position, age, education, gender and ethnicity.

Section 1 collects information on the population and geographic region of the commune. It also collects information on land area, number of households, and number of agricultural households, the main ethnicities, religions of the population, and whether the commune was listed in the 135 program or considered as remote area by regulation of the government.

Section 2 contains information on major sources of income in the commune, the changes of living standard and reasons for the changes in standard of living of the people of the commune, types of aid programs received in the past 12 months and from whom, types of *ad hoc* aid, the project or programs the government or other organizations are currently implementing in the commune such as job creation, poverty alleviation, infrastructure, etc., the number of households considered as poors by the MOLISA method, number of poor households receiving various types of assistance such as subsidized credit, exemption from or reduction of school fee, relief fund for natural disasters and pre-harvested shortage, occupational and agricultural technology training, exemption from production tax.

Section 3 collects information on the existence, the number of enterprises/factories, cottage industries or handicrafts operating within 10 km from the commune center, types of enterprises, number of enterprises having more than 20 employees, number of enterprise located in the commune, the traditional businesses existing in the commune, the ownership of these traditional businesses, the time that enterprise was set up, number of employees, and the average wage for these employees.

Section 4 asks questions about the land distribution of the commune, main crops grown in the area, number of crops harvested each year and whether the productivity of the crops increased or decreased during the past 12 months and the reasons for that. This section also collects information about types of land of the commune, total area for each type of land, percentage of irrigation, quality for annual planted land area, and aquaculture water surface land. Daily wage rates for various agricultural works for men, women and children under 15 are collected. The existence and distance of an agricultural extension center, and the number of visits of extension workers are also asked about. This section also asks questions about the main problems faced by farmers.

Section 5 collects information about roads and waterways at the village level. The questions include the distance from the village to the nearest road for cars, number of months per year the road was not used by cars, the type of road surface of the road coming to the village, the same information for waterway, and the distance and time from the surveyed village to various places such as the commune people's committee building, district center, post-office, telephone, food store and market. Questions are also asked about the main source of drinking water for most of people in the commune in the dry and rainy seasons, the availability of electricity in the commune and the frequency of electricity outages.

Section 6 collects information on distance, method of transportation and time required to go to the nearest schools of each level from the surveyed village. For each level of schooling, the reasons why children dropped out and the main education obstacles of the commune are asked about. Information on anti-literacy programs, the starting time and number of participants by gender are also collected. For the surveyed village, information on preschool is collected.

Section 7 gathers information on the main illness/diseases of concern in the commune, the use of health services, major problems with the health services in the commune and distance and time required to go from the commune to each of various health facilities.
Section 8 gathers information on the most important social issues in the commune such as drugs addiction, prostitution, gambling, drunk, theft, superstition, etc., the number of people in the commune related to these issues and the number of which are children under 16 years old, and the number of people who receives regular social assistance from the commune. 2.2. The VHLSS 2004 Questionnaires

The VHLSS 2004 questionnaires are developed based on the VHLSS 2002 questionnaires to ensure the comparability between the two surveys. There are some changes in the household and commune questionnaires that will be discussed in the following sections. The major changes in the VHLSS 2004 household questionnaire are two additional new modules for the long household questionnaire in section 9 (additional section) and 10 (non-farm self employment activities). This section will mainly discuss the changes in the household and commune questionnaires in 2004 compared to the questionnaires in 2002.

## 2.2.1. The Household Questionnaire

## SECTION 0: COVER PAGE

The main difference in the cover page of the household questionnaire in 2004 compared to 2002 is the information to define whether the household is surveyed in 2002 or not. That section includes the code of province, district, commune/ward, enumeration area, household and quarter of interview in 2002.

### SECTION 1: HOUSEHOLD ROSTER

There are some changes in this section. The section is divided into two parts. The first part (part A) has some changes. The question asking about relationship with the household head combines all kids of the household head into one code, and natural brothers and sisters with other relation into another code. The date of birth for each member in the household excluded the information about the day of birth. In this section, there are three additional questions to collect information about number of months that the household member had actually been living in the household in the past 12 months, the registration status of each household member and number of years and months the household member had been living in the current city/province. The second part (part B) is a new part to collect information about household members of the household that was surveyed in 2002, including individual code, gender, age in 2002, currently living in the household or not, the individual code in 2004 and the reason for moving out of the household.

### **SECTION 2: EDUCATION**

There are a lot of changes in this section compared to 2002. The question asking about the highest diploma the household member received is coded in more detail and divided into two columns (one for education and other for professional training). There is one more answer option (school does not ask for fee) in the question asking about the reason for school fee and contribution exemption and this question is divided into two columns (one for school fee and one for contribution). There are three additional questions to collect information about whether the household member currently goes to school or not, the grade and types of school (public or private) of the household member, if any and the amount received from social organization.

# SECTION 3: HEALTH

In 2002, section 3 is the employment section but it is moved into section 4 in 2004. Section 3 collects additional information compared to 2002 about whether any household member was sick or ill during

the past 4 weeks before asking about her or his health status in the past 12 months, number of days in bed with the help of other people, number of days the household member could not work as normal, whether she or he has health insurance and receives from whom, which health facility she or he uses and the reason to use that health facility. The questions asking about the cost of using health services are collected in more detail for each time of visit for inpatient and outpatient treatment. In addition, this section also collects information about whether or not the household has enough money to pay for health costs and if not, how they resolved this problem.

## **SECTION 4: INCOME**

This section of VHLSS 2004 combines sections 3 and 5 in VHLSS 2002. The part 4A includes information of section 3 and section 5A of VHLSS 2002.

In section 4 part A, the question "did you look for a job or would like to work in the past 7 days" is omitted. However, this section collects in more details information about the main job and the second job in the past 12 months, including the number of working months, number of working days per month and the number of working hours per day, the number of working years. There are also additional questions to know whether the household member is a government official or not, the total amount received from the third job, total hours of housework in the past 12 months.

Section 5B1 in VHLSS 2002 is moved to section 9 of VHLSS 2004. There are major changes in questions about agricultural production (section 4B) in VHLSS 2004. Rice production is separated out from other crop production,. In addition, questions about the rice crops are more detailed than in 2002. It collects more information about the total loss in the past 12 moths, the amount used as seed for next crop, the amount used for raising livestock, the amount used as gift and the rest. There is an additional question to collect information about the amount the household uses for consumption in other annual crops, annual and perennial industry crops, and fruit but it does not collect information that to whom the household sold these products in 2004. Instead of collecting total value for each by-product of crops as in 2002, it collects information about the value sold, the value of feeding livestock and the value using for other household purposes in the past 12 months for each crop by-product.

The main differences in the crop expenditure compared to VHLSS 2002 are the more detail in cost items and the separated cost of rice.

The section on livestock production (5B3 in VHLSS 2002 and 4B2 in VHLSS 2004) has more details about the livestock than in 2002 in both revenue and cost. However, there is no question about the cost of maintenance and small repair in 2004.

In the section "forestry" (section 5B5 in VHLSS 2002 and section 4B4 in VHLSS 2004), the information for four activities (forest plantation, protection, maintenance, improvement; trees for breeding and other products collected from forests; other forestry services; and hunting, trapping, domesticating forestry animals and birds) are collected in more details, including the value of sales and the total value in 2004. The cost of each item in this section is collected separately for each main activity (forestry production, forestry service and hunting activities) in 2004 instead of asking for all as in 2002.

There is a minor change in the cost of aquaculture production. The cost of medicines is combined with other expenses in 2004. The non-farm self employment activities (section 5C in 2002 and 4C in 2004) are designed to collect additional information about these activities of the household in 2004. The code of the household member that knows most about this activity and the value of the product currently in stock are omitted. The number of operating days per operating month, the ownership status of the activity, the percentage of income from the activity the household received, the value of product from this activity is exchanged, and information about the value of by-products from this activity are added in 2004. The cost of each item is asked for each operating month instead of asking for the whole year as in 2002. In addition, the cost items are divided into more details (14 cost items) than in 2002 (10 cost items).

The main difference in the remaining part of the income section (section 5D in 2002 and 4D in 2004) is that "the income from charity and support from donor or organization and others" item is divided into two separate items that are "income received from charity and support from donor or organization", and "income received from others".

# SECTION 5: EXPENDITURE

This section is section 6 in 2002. The main difference in section 5A (6A in 2002), is that the food item (chewing gum) is dropped in 2004. In section 5B2 (section 6B2 in 2002), the expenditure on vacation is divided into 2 categories, namely expenditure on domestic and overseas vacation.

# SECTION 6: FIXED ASSETS AND DURABLE GOODS

This section is section 7 in VHLSS 2002 and it is divided into two subsections (section 6A for fixed assets and section 6B for durable goods). In addition, the telephone is divided into mobile phone and fixed phone and there are two more items than collected in 2004 (micro wave ovens and fruit grinding machines). Three more questions are added in section 6A. The quantity of each asset, the percentage the household owned, and the total amount the household spent on repair and maintenance for all assets during the past 12 months. In section 6B, the question about quantity of the asset is added.

# SECTION 7: HOUSING

This is section 8 in VHLSS 2002. There are many changes such as omission and addition of more questions in this section compared to 2002 questionnaire. It omits the questions about previous place the household lived, the material made for toilet cover, and all the questions (from question 34 to question 44) that collect information about accessing public information such as television and radio channels, and newspapers. It adds more questions about the number of months the household has been renting the house in the past 12 months, the number of months the household has been renting out the land or house in the past 12 months. It divides the water source into drinking water and water for other uses and collects information about how frequently the household boils the drinking water.

# SECTION 8: PARTICIPATION IN POVERTY REDUCTION PROGRAMS

This is section 9 in VHLSS 2002. This is nearly new content in this section compared to the 2002 questionnaire. The first question defines whether the household was classified as poor in 1999 or not and then it collects information about whether the household knows about the National Target Program and the 135 Program, and through which information channel (television, radio, newspaper, etc.), the main projects from these programs, the benefit that the household got from these programs. The next questions collect information about the evaluation of the changes in their living, the reason for those changes, and the difficulties that they met in doing their business in the past 12 months. The information about the borrowing of the household is collected for both formal and informal lenders (there is only information about formal lenders in 2002) with further details about the date, using purpose, amount paid for both principal and dividend during past 12 months, and the intended date to pay back for each loan.

### SECTION 9: AGRICULTURE, FORESTRY AND AQUICULTURE PROUCTION (ADDITIONAL MODULE)

- 9.1 Agriculture, forestry, and water- surface, residential land and garden, pond next to housing land
- 9.2 Land that is rented/borrowed/temporarily exchanged

9.3 Land that is temporary rented out/lent out

9.4 Planting

9.4.1 Rice

9.4.2 Staple food, food crops and other annual crops

9.4.3 Annual and perennial industrial crops

9.4.4 Fruit crops

9.4.5 Cropping structure

9.4.6 Access to extension services

9.4.7 Conversions in agricultural, forestry and aquaculture land and over the past 10 years

9.4.8 Bought, bid, inherited land or use right-transferred land over the last 10 years

9.4.9 Sold, tender-expired, acquired land or inheritance right-transferred land in the last 10 years.

Section 9.1, 9.2 and 9.3 is the section 5B1 in VHLSS 2002.

Section 9.1 collects additional information about the number of plots, the distance from the household residential to each plot, the quality, irrigating method, the ownership status, the first time the household used, the year the household has Land Use Certificate, the main household member using and managing and whether the household uses the plot for production purpose during the past 12 months.

Section 9.2 collects in more detail information about rented/borrowed/temporarily exchanged land of the household in the past 12 months. It collects information about the relationship with the landlord, the duration, payment method, the amount of money the household paid, for how many months, and the date of the latest payment.

Section 9.3 collects information about land the household rented out in the past 12 months and it collects the same information as in section 9.2.

Sections 9.4.1 to 9.4.4 are mentioned in section 5B2 in VHLSS 2002. However, they are divided in further details for rice and other crops. These sections mainly collect information about productivity of rice production compared to the previous year and the reason for the change in the productivity of rice production and who is the main trader in the past 12 months. For other crops, the main information these sections collected are about the main traders and the place the household sold the products.

Section 9.4.5 collects information about the area of each type of crop the household planted in 1993, 1998 and 2003. In addition, it collects information about, the change in rice production (variety) from 1998 up to the interview time, the year using these varieties, the year the household stopped using the varieties, from which information source the household knows about the variety, who they brought from and the reason they are using it. The next questions collect information about the new crops the household planted since 1994, the first planted year, the information source about the new crops, whether the household is still planting it or not, and the reason as well as the change in planted area for these crops.

Section 9.4.6 collects information about accessing extension services. It is a new section in VHLSS 2004. It collects information about 15 extension services, including fertilizer, irrigation new seeds, land issues, weather and livestock service, food for livestock, market information, etc. In addition, information about the fee and the impact of these services on decision making of the household on crop and livestock production is also collected.

Section 9.4.7 collects information about conversions in agricultural, forestry and aquaculture land during the past 10 years by type of land before and after conversion, area of conversion, the time and cost of conversion, amount received from government assistance in both cash and kind for these conversions. In addition, this section collects information about forestry planting including the area of planting, the first and the last year of planting, the cost and amount of cash and kind received from government for each plot.

#### MARKET, POLICY AND POVERTY REDUCTION IN VIETNAM

Section 9.4.8 collects information about land the household bought, bid, inherited or use right-transferred over the last 10 years, including month and year the household owned, month and year the household received land certificate, type of land, area, the reason the household has this land, the cost in both cash and kind and from whom the household got the land.

Section 9.4.9 collects information about land the household sold, tender-expired, acquired land or inheritance right-transferred land in the past 10 years. The same questions as in section 9.4.8 are used to ask the household such as the time they were sold; area, amount received, and type of land, etc. However, there are some additional questions about the reason and the compensation received for land that was acquired by government, the money received and the relationship of the buyer for the land sold.

### SECTION 10: NON-FARM ACTIVITIES (ADDITIONAL MODULE)

- 10A Information about time, location and labor
- 10B Business history
- 10C Participation in business associations and clubs
- 10D Contact with government agencies
- 10E Other characteristics
- 10F Activities that operated during the past 10 years and stopped in the past 12 months

Section 10A collects information about whether the non-farm activities are operated in every month in the past 12 months or not and the reason for not operating during the whole year, the operating place, the total number of labor and number of paid labor.

Section 10B collects information about the history of the non-farm activities, including the opening year, the founder, number of other competitors, and the relationship of the household with these competitors, the main difficulties at the beginning of this business, total amount of money invested at the beginning, from which source the household has this amount of money, experience about this business before setting up.

Section 10C collects information about whether the household is a member of Vietnam Chamber of Commerce and Industries (VCCI), member of other business union, enterprise union, and the services the household received from these unions.

Section 10D collects information about the number of times tax collectors visit the household and the reason for these visits.

Section 10E collects information about the way the household buys inputs for these non-farm activities and the place the household sells the outputs of the activities as well as evaluation of the importance of each market for these products (internal province and outside market as well as international market). This section also collects information about the information channels the household often use in order to set the price for its products, information about the business performance compared to two years ago as well as the major constraint for these household businesses.

Section 10F is the last section of the household questionnaire. It collects information about the nonfarm activities the household had operated during the past 10 years but stopped to operate during the past 12 months, including the first year it started, number of operating years, the household member who operated it and the reason for stopping these activities.

#### 2.2.2. The Commune Questionnaire

There is an additional section (section 9) in the VHLSS 2004 commune questionnaire compared to 2002.

This section collects information about credit and saving, which will be discussed in more details later. We will only describe the changes between the VHLSS 2004 and 2002 commune questionnaires in each section.

Section 0: Two questions are added in this section to collect information about the number of years living in this commune of and previous position that held by each local commune leader.

Section 1: This section drops the questions about the year of data collected, the number of people working in the agriculture sector and adds more questions about the number of households and people living in the commune by registration status and the number of people migrating in and out of the commune during the past 12 months.

Section 2: This section collects additional information about the number of households affected by natural disasters in the past 3 years, the number of households classified as a poor household by the National Target Program and the amount of money received as assistance for natural disaster relief during the past 3 years.

Section 3: This section drops the questions about total the number of employees, the number of employees from each of the 5 main factories/enterprises and if traditional businesses increased or decreased compared to 5 years ago and the average labor cost. It adds additional questions about the number of laborers in the commune and the number of female laborers of the commune that are working for these enterprises. In addition, it also collects further information about enterprises/traditional businesses that laborers of the commune could go to work and come back home daily, the number of enterprises/traditional businesses located within the commune and the name of traditional businesses.

Section 4: In VHLSS 2002, this section is collected for both urban and rural communes in which agriculture is the main income source. However, it only collects information for every commune in the rural areas in 2004. In addition, this section collects further information about the productivity of crop production compared to 5 years ago, the reason for the change, the total crop output compared to 5 years ago, where the farmers sold their agricultural products, the main trader, the distance from the commune to the market, the assistance received from the trader, the number of traders to whom the farmers in the commune could sell their products, the types of crops and the total area by crop in 2003, 1998 and 1993. Information about the land situation in the commune is also collected, including the irrigation by type of land, percentage of land having Land Certificate (LUC), the reason for the land having no LUC, the year that the biggest LUCs are provided, the number of LUCs in that year, the main reason the household would like to have LUC. There is also additional information about the free time and lack of work for labor by gender each month in the past 12 months . In the questions asking about extension service, there are additional information about the percentage of farmers who participated in the extension meetings, the percentage of women who participated in these meetings, the services that these meetings could not provide, the number of times that an extension official visited the farmer and whether the commune has a veterinary official and the most important difficulties the farmer faced.

Section 5: This section collects additional information about the radio station, cultural communal house, and small irrigation station in the commune. It also collects additional information about the main mean of transportation, the owner status of the transportation mean, cost, time, and the frequency that people living in the surveyed village often use to go to market, committee house, post office, bank, small town, center of province, big cities such as Hanoi, Haiphong and Ho Chi Minh City. It also collects more information about the main goods that is sold in the daily market, main market, and the projects involving infrastructure that are implemented during the past 10 years, including the type of project, starting year, completion year, main budget sources, the value of the project and the contribution of the commune comprising household contributions, the number of benefiting households in the commune and the number of surveyed households that benefited from the project.

Section 6: There are only two more questions added in this section in 2004 asking about whether the commune has a kindergarten or preschool before asking in detail information about these schools in the surveyed village as in 2002.

Section 7: There is only one additional question asking about the transportation mean that people living in the commune often use to go to health facilities.

Section 8: There is no change in this section.

Section 9: This is a new section in the commune questionnaire. It collects information about credit and saving. First, it collects information about the main saving methods that people in the commune often use and the place the people can place their money by type of organizations (bank, credit institution, union, private), the distance from the commune to these organizations and whether the households in the commune can borrow from these agencies, and the purpose of the loan.

### 3. The Sample

#### 3.1. Sample Design

The series of VHLSSs from 2002 to 2010 rely on a master sample for sample selection. The master sample is a random sample of the 1999 Population Census enumeration areas. From this sample of enumeration areas, multiple samples of households can be selected for different surveys or for the different years of a rotating panel survey such as the VHLSS. The master sample used in the VHLSS is a two-stage area sample where communes are selected in the first stage and 3 enumeration areas (EA) per commune are selected in the second stage. The communes were stratified on province and urban/rural and the sample was allocated over strata proportional to the square root of the number of households. Both communes and EAs are selected with probability proportionate to size (PPS), with the size being the number of households according to Population Census 1999. This section describes in detail the master sample and other aspects of the sample design established for the VHLSS.

#### 3.1.1. Target Population

The target population of the VHLSS comprises the civilian, non-institutionalized population of Vietnam. In order to cover this target population, interviews are conducted at the household level. Because people are mobile, it is important to define clearly where people are to be enumerated to avoid double counting and under coverage. Only persons considered as permanent residents of the household are eligible for inclusion in the surveys. Persons who are considered as permanent residents of a household but are away temporarily will be included. This includes persons on vacation, temporarily in a hospital, and students living away from home during the school year. However, any housing units containing only students living away from home during the school year should not be included in the survey because data for such students would be obtained from their permanent place of residence.

*Ineligible households.* GSO has to decide on eligibility criteria for the VHLSS. Some households may be considered as ineligible for selection in the survey. One example is student housing blocks. The households consist of rooms in a dormitory, and not actual family households. They are unstable, changing substantially from year to year. The same may be true for households in other institutions (military compounds, hospitals, prisons, etc.).

*Ineligible EAs.* Some EAs consists entirely or to a large part of households that are not eligible for selection in the survey. EAs which consist primarily or entirely of ineligible households should, if possible, be excluded from the sample frame. For example EAs that consists almost entirely of student housing blocks.

#### 3.1.2. Three-Stage Design

The sample design consists of three stages with communes/wards selected at the first stage, census enumeration areas (EAs) as the secondary sampling units (SSUs) and households selected at the third stage.

#### PRIMARY SAMPLING UNITS

The sampling units at the first stage of the sample selection are called primary sampling units (PSUs). PSUs must be sufficiently large to allow for repeated sampling of households within them. In Vietnam there are two choices of PSUs for a master sample: communes (or their urban equivalent the district town or ward) or EAs. Communes contain on average about 1,600 households while EAs were set up during the census to contain about 100 households (Table 1). PSUs should contain a minimum measure of size of 70 to 75 households to permit repeated sampling of households for the VHLSS or other surveys. However, about 8,000 EAs (5 percent of all EAs) containing less than 70 households would have to be combined, which would be a time-consuming task.

Based on the above considerations, communes were chosen as PSUs, with three EAs to be selected per commune at the second stage, using only one of EA per survey. This is technically a three-stage design (counting the selection of households), but it is operationally equivalent to a two-stage design. This design solves several problems simultaneously:

• The problem of the large size of communes does not arise because we are working only one of the selected EAs for each survey.

• This design allows for rotation of EAs, rather than households, which is operationally less complicated.<sup>134</sup>

• The problem of small EAs is not as serious since it is the EAs, rather than households, that will be rotated over time. If the need to group EAs to satisfy a minimum measure of size arises, then such grouping can be restricted to only the selected communes.

Communes as PSUs should also not be too small. A minimum size of 300 households is required for the commune to serve as a PSU because a sub-sample of three EAs is to be selected from the commune. Altogether 529 communes/wards in 38 provinces were found to have less than 300 households. Some provinces have a large proportion of small communes, for example rural Lao Cai where 50 out of 161 communes have less than 300 households. Only 4 urban wards have less than 300 households.

	Urban areas	Rural areas	All areas
Number of households (1000)	4,026	12,635	16,661
	(24%)	(76%)	(100%)
Number of EAs	38,435	128,085	166,520
	(23%)	(77%)	(100%)
Number of communes/wards	1,561	8,915	10,476
	(15%)	(85%)	(100%)
Households per EA	105	99	100
Households per commune/ward	2,579	1,417	1,590
Number of EAs per commune/ward	25	14	16
Source: Census of Population and Housing 1999.			

Table 1: Summary Statistics on Communes/Wards and Enumeration Areas, 1999

<sup>134</sup> As explained below, each time the VHLSS is completed, half of the households will be replaced by new households. This replacement is referred to as rotation.

# SECONDARY SAMPLING UNITS

Because some communes/wards are quite large, a two-stage design with selection of households randomly within a commune/ward would have created difficulties in implementing fieldwork, so a second stage of selection was introduced in which three EAs are selected per commune. However, in each survey only one of the three EAs will be used. This design allows for rotation of EAs, rather than households, which is operationally simpler. Also, the problem of small EAs is not as serious since it is the EAs, rather than households, that will be rotated over time.

### THIRD STAGE SAMPLING UNITS

Households are the third stage sampling units. Sample size in terms of number of households and the number of households to be selected per EA will be discussed below.

### 3.1.3. Sample Size

The targeted sample size for the VHLSS 2002 was 30,000 completed household interviews in order to get precise estimates at the provincial level. A much smaller sample size would have been adequate for the production of precise estimates at the national and regional levels and for urban and rural areas within these domains.

#### 3.1.4. Cluster Size

There are two main alternatives for selecting cluster size. If the cluster size is allowed to vary, the number of households per cluster can be selected so that the overall probability of selection within a province is the same for all households. The alternative is to select a fixed number of households with sampled EAs. Selecting a fixed number of households per EA is easier to implement and is the procedure traditionally used by the GSO. However, this procedure produces more variation in the sampling weights, which reduces the precision of survey estimates.

For the VHLSS, an optimal cluster size would be 10 in urban areas and 20 in rural areas, but because of high costs of implementing the survey in a large number of clusters, the final cluster size was selected to be 25. EAs with minimum size of 70 households are of adequate size for selection of 25 households per cluster.

### 3.1.5. Augmentation of the Sample in 2002

After design and selection of the master sample was completed, there remained a strong concern that the sample was too small and would only produce an estimated percentage at the provincial level with a confidence interval of plus or minus six percentage points. After much discussion the GSO decided to augment the sample for the VHLSS 2002 in all provinces, with the augmented sample only responding to a shortened version of the questionnaire focused on collecting income information.

The augmentation of the sample was done by doubling the number of PSUs to 700 urban wards and district towns and 2300 rural communes. In 2002, the original sample was divided into 4 parts implemented in one month of each quarter of the year (see Table 2).

#### Table 2: Distribution of the Sample in VHLSS 2002 by Quarters

Quarte	r I Quarter I	I Quarter III	Quarter IV	Total

Long questionnaire	7,500	7,500	7,500	7,500	30,000
Short questionnaire	22,500	22,500			45,000
Total	30,000	30,000	7,500	7,500	75,000

#### 3.1.6. Stratification

The purpose of stratification is to increase the precision of survey estimates. Prior to sample selection, the sampling frame needs to be divided into mutually exclusive and collectively exhaustive categories called strata. Sample selection is then carried out independently within the strata. The formation of the strata depends to a large extent on the domains of interest for the survey. In response to the demand for accurate data at the provincial level the master sample used explicit strata based on the provinces and urban/rural status within provinces.

There is a large variation in the number of households among provinces from 57,000 in Bac Kan province to over 1 million in HCMC. This leads to a conflict between demands for equal precision in estimates for all provinces (implying equal sample sizes within provinces) and the interest to get as good precision as possible in the national estimates (implying sample sizes proportional to the population size in the provinces). A compromise solution was used in the VHLSS by selecting the sample size in each provincial stratum proportional to the square root of the number of households in the given stratum.

Within each province, further stratification was done by urban/rural residence to ensure high level of precision in the urban and rural estimates. The number of urban and rural PSUs was allocated proportional to the number of households in urban and rural areas.

#### 3.1.7. Sample Rotation

The procedure for achieving a fifty percent rotation of households from one implementation of the survey to the next (every two years for the VHLSS) depends on the units used to construct the master sample. With the master sample PSU being the commune, the fifty percent rotation is implemented by retaining the EAs in half of the communes and selecting new EAs in the other half.

### 3.2. Sampling Procedures

Sample selection of the communes/wards and EAs was done as part of the development of the master sample. Each year the survey is implemented, one of the EAs in each commune must be selected, the household listing is updated in that EA and the sample of survey households is selected. Sample selection basically consists of preparing and updating a sample frame and randomly selecting units from that frame.

#### 3.2.1. Selection of the PSU

The sampling frame for selection of enumeration areas in the Household Living Standards Survey is the list of all enumeration areas identified and used in the 1999 population and housing census of Vietnam.

After examining the Excel sheets listing commune names and sample size kept by the Labor and Population Department of the GSO, about 529 communes in 38 provinces were identified to have less than the minimum measure of size (300 households) required for a PSU. All the communes with less than 300 households were combined with adjacent communes using administrative maps to ensure contiguity.

Within each province, the name and measure of size (number of households from the 1999 Population Census) in urban wards and district towns were listed separately from rural communes to allow independent sample selection within urban and rural strata.

The sampling of primary sampling units was done with systematic PPS sampling separately within

each stratum (province\*urban/rural) by the Social and Environmental Statistics Department of the GSO. The communes (or combined communes) within each province are listed in order from north to south and east to west ensuring a geographical spread of the sample. The same was done in the urban stratum although in this case the listing was done within each urban location.

The next step was to compile the cumulative number of households for the PSUs on the list. The total number of households in the stratum is then divided by the number of PSUs to be selected from the stratum. The result, *s*, is the "step" to be used as sampling interval in the systematic selection. Next, a random number, *a*, between 1 and *s*, was selected. The first PSU selected is the PSU containing household number *a*. The next PSUs to be selected are the PSUs containing household number a + s, a + 2s, a + 3s and so on.

The result of the original sampling was 700 urban wards and district towns and 2300 rural communes.

#### 3.2.2. Selection of Secondary Sampling Units

During census data collection, communes or wards were partitioned into Enumeration Areas. Maps of relatively high quality exist for all communes/wards, and all EAs within each commune/ward. The commune/ward maps clearly identify the EAs located in them. For the selected communes/wards a list of all the EAs was prepared. All EAs with less than 70 households were combined with adjacent EAs. (With some adjustments in remote, mountainous areas). Lists of EAs were obtained from CD-ROM discs at the Department of Population and Labor.

Within each selected commune/ward three EAs were selected. The selection was done by a PPS procedure in the same way as was done when selecting communes/wards within the stratum. Numbers 1, 2 and 3 were assigned to the selected EAs according to the selection sequence. All EAs having number 1 (one EA in each selected commune/ward) constitute the sample for the first survey year (2002).

#### 3.2.3. Selection of Households

The sampling frame for selection of households is a list of all households in the enumeration areas updated by provincial statistical officers visiting the communes/wards during the year prior to implementing the survey.

In the EAs a sample of households was selected systematically with 20 households in rural EAs and 10 households in urban EAs. The selection procedure was as follows: The total number of households in the EA was divided by the total number of households to be selected from the EA. The result, *s*, is the step to be used as sampling interval in the systematic selection. Next, a random number, *a*, between 1 and *s*, was selected. The random number was selected by the field supervisor from a list of random numbers provided by GSO. The first household to be selected is the household number *a*. The next households to be selected are the household number a + s, a + 2s, a + 3s and so on.

#### 3.2.4. Sample Selection for the Augmented Sample for 2002

In the urban domain 700 EAs were selected with a sample size of 25 per EA. We have selected 175 EAs and 25 households/EA for the ordinary survey in January. For the other three survey months in 2002 we have selected 525 EAs and 10 households/EA. The additional sample consists of an additional 10 households in each of the 175+525 urban EAs. This gives an additional sample of 7,000 urban households.

In the rural domain we have selected 287 (or 288) EAs and 20 households/EA for the ordinary survey in January. For the other three survey months in 2002 we have selected 863 (or 862) EAs and 20 households/EA (altogether 1150 EAs for the four survey months). It was not possible to use the

ordinary EAs for the additional sample as was done for the urban domain. This would result in a sample of 40 households from each EA, which is far above the optimum. Instead, we selected a *parallel sample* of 1,150 PSUs and one EA in each selected PSU. This was done by repeating the steps for selecting the ordinary rural sample. There was in this case no need for ensuring an EA size of at least 70 households (as we had to do in the ordinary sample), it sufficed to ensure a size of at least 25 households.

# 3.3. Revisions to Sample Design for the 2004 VHLSS

The implementation of the 2002 VHLSS provided valuable information which was useful in reducing sample size for the 2004 VHLSS.

#### 3.3.1. Changes to the Sample Size for the VHLSS 2004

Analyses of the 2002 survey results indicate that a sample with the ability to detect rather small short term changes at the provincial level must be very large, so large that it would be very difficult to manage. A sample of, say, 3-4,000 households in the province would require hiring of temporary staff which is likely to create problems with the quality of the field work and the quality of the data, hence reducing the value of the survey.

In the 2002 survey the sample taken within enumeration areas (cluster size) was 25 households. Calculations show that there is very little information gain from the last households in the sample from the enumeration area. Based on these findings, the VHLSS 2004 survey sample design was revised as follows:

• Cluster sizes were reduced to 15 households per enumeration area without any expected serious effects on the precision.

• The sample size of the expenditure module was reduced to 9,000 households allowing comparisons over major groups of households and individuals, but not comparison at the provincial level.

• The sample size of the income module was reduced to 36,000 households by reducing the number of households sampled per PSU from 25 to 15, not by reducing the number of PSUs.

• The proposed reductions in sample size only marginally reduced the probability of detecting changes in the poverty rates as the ability to detect small changes at the provincial level were already very small with the larger 2002 sample size.

• For the 2004 survey altogether 3,063 EAs were included in the sample. Some 1,546 EAs were selected as a sub-sample from the VHLSS 2002 sample. The remaining 1,517 EAs were selected with systematic PPS from 1,517 communes. Additional EAs were selected in 2004 for 6 new provinces so that the total number of EAs is more than 3,000.

• For the "old" EAs (sub-sample from the VHLSS 2002 sample) the households were selected as a sub-sample of the VHLSS 2002 households.

• For the "new" EAs the following procedure was used: Within selected EAs all households were listed and 20 households were selected with systematic equal probability sampling. A sub-sample of 15 households was selected systematically from the 20 households. This sub-sample was used for the survey and the remaining 5 households were designated as reserve households (to be used as substitutes for non-response households). From the 15 households a sample of 3 households were selected randomly by systematic sampling. This sub-sample was used for the expenditure module of the survey.

#### 3.3.2 Differences Between the 2002 and 2004 Sample Designs

The sample design for the 2004 survey is in many ways better than the design for the 2002 survey. The cluster sizes for the expenditure module are smaller, 3 households per EA as compared to 5 or 20

households per EA in the 2002 survey. The design effect due to clustering will consequently be smaller in the 2004 survey.

The selection of a second stage sample in EAs of either 5 or 20 households in the 2002 survey resulted in a large variation in the sampling weights. The variation in the weights increases the variances as compared to a self-weighting design. The increase in the variances due to variation in the weights can be measured by the design effect due to weighting.

Table 3 shows the average design effects for the two surveys. The design effect could be looked upon as the "cost" in terms of increased variance we have from the design (as compared to a simple random sampling design). It is apparent that the "cost" for variation in the weights in the 2002 survey is high. This is due to the varying sub-sample sizes (either 5 or 20 households) in the 2002 survey.

Table 3: Average Design Effects due to Weighting and Clustering in the Two Surveys

	Design effect due to:		
	Weighting	Clustering	Total
VHLSS 2002	1.85	1.78	3.29
VHLSS 2004	1.12	1.2	1.34

The *effective* sample size is defined as the number of sample observations divided by the design effect. The effective sample sizes for the two surveys (expenditure module) are:

2002 survey: 29,530/3.29 = 8,975

2004 survey: 9,189/1.34 = 6,857

The effective sample size is the "real value" of the sample in terms of a simple random sample. It is interesting to note that if the 2002 survey had used sub-sampling of an equal number of households in the EA (as the 2004 survey) the design effect would have been  $1.12 \times 1.78 = 1.99$  and the effective sample size would have been = 14,839.

From the effective sample sizes we can calculate the expected ratio between coefficients of variation (CV) for the two surveys:

$$\frac{CV_{2004}}{CV_{2002}} = \sqrt{\frac{8975}{6857}} = 1.14$$

This shows that the sampling errors in VHLSS 2004 are expected to be only 14 percent above the sampling errors for the VHLSS 2002. This is a very small difference considering that the sample size in 2004 was less than one third of the 2002 sample.

### 3.4. Sampling Evaluation

#### 3.4.1. Change in Definition of Households

The "demarcation" of households is another problem. There is a slight difference between the 2002 and 2004 questionnaires. In the 2004 survey the questions on household members are followed by a question on how many months each person has been staying in the dwelling during the last 12 months and a question on registering for residency. In the 2002 survey these two questions were not asked. Errors in household sizes affect poverty calculations. Crude calculations show that if the average household size contains a 5 percent under-reporting error the poverty rate will be biased downwards by approximately 3 percentage units.

#### 3.4.2 Comparability of Sample Population Estimates to Population projections

Table 4 shows estimated number of households and population from the surveys 2002 and 2004 and population projections based on census data from 1999.

Year	Number of households (million)	Population (million)	Population according to population projections (constant fertility variant)
1999	16.7	76.3	76.3
2002	17.28	76.6	79.7
	(0.02)	(0.34)	
2004	17.34	75.5	81.5
	(0.02)	(0.23)	

Table 4: Comparison of Survey Estimates and Population Projections

Notes: Standard errors in parentheses.

The total number of households and the population in 2004 seems low. The annual growth of the population during the period is 1.3 percent according to the population projections (1.2 percent with the alternative "declining fertility rates"). Assuming a slowly declining average household size we should expect the growth rate of households to exceed that rate somewhat. The survey, however, shows an annual growth rate of 0.8 percent between 1999 and 2004. There is possibly an under-listing of households in the selected EAs.

The estimates of total population are low compared to the census projections for both surveys but especially for the 2004 survey. This is partly due to the fact that the target population was confined to civilian persons residing in households, thus excluding persons living in institutions and military personnel living in barracks. There may also be effects of either too low household estimates or under-enumeration of household members or both.

### 3.5. Weights for the VHLSS Samples

### 3.5.1. Calculation of Sampling Weights

In order to facilitate the calculation of sampling weights, an Excel file was created which recorded the relevant information and calculated the probabilities of selection at each stage. From this spreadsheet the sampling weights were calculated.

First stage sampling weight:

 $w_{1hi} = \frac{1}{n_h} \frac{A_h}{A_{hi}}$   $n_h = \text{number of PSUs selected in stratum } h$   $A_h = \text{number of households in stratum } h \text{ according to census}$   $A_{hi} = \text{number of households in PSU } i \text{ in stratum } h \text{ according to census}$ 

Second stage sampling weight:

 $w_{2hij} = \frac{A_{hi}}{A_{hij}}$ 

 $A_{hi}$  = number of households in PSU *i* in stratum *h* according to census  $A_{hij}$  = number of households in EA *j* in PSU *i* in stratum *h* according to census Third stage sampling weight:

 $w_{3hij} = \frac{M_{hij}}{m_{hij}}$ 

 $M_{hij}$  = number of households in EA *j* in PSU *i* in stratum h according to survey listing  $m_{hij}$  = number of households in the sample from EA *j* in PSU *i* in stratum *h*. The final weights, to be applied to each household in the data set, become:

$$w_{hij} = \frac{1}{n_h} \cdot \frac{A_h}{A_{hi}} \cdot \frac{A_{hi}}{A_{hij}} \cdot \frac{M_{hij}}{m_{hij}}$$

First stage weights for approximately 100 communes are less than 1.00. This happens when the measure of size for the commune (number of households) exceeds the sampling interval (the number of households in the stratum divided by the number of communes selected in the stratum).

The best way to handle these large communes (larger than  $M_h/n_h$ ) would have been to set them aside in a special stratum before selection and to select all communes in this stratum with certainty (probability = 1).

Some of the large communes have a first stage weight that is less than 0.5. In that case the commune will be selected twice when the sampling interval is applied. This will give a sample of two EAs in the commune in the second stage selection. It appears, however, that in these large communes only one EA has been selected. Whether this was deliberately done or a result of error in the sample selection has not been investigated.

Measures of size from the 1999 Census are missing for 62 communes in the file weights04. These are cases where the commune has been formed by a split from one or several communes since 1999. The missing values mean that a proper PPS selection of these communes could not be done. The best procedure would have been to distribute the 1999 census number in the old commune(s) between the new communes as good as possible in order to get a measure of size of reasonable quality also for the new communes. This seems not to have been done.

#### 3.5.2. Sampling Weights for VHLSS 2002

Figure 1 shows the sampling weights for the 30,000 households interviewed for the expenditure module in the 2002 survey. Weights above 2,000 have been excluded in the diagram (315 households had weights above 2,000).

The bimodal appearance of the histogram is due to the two alternatives of sub-sampling of households in the EAs (5 households or 20 households). It is obvious from the histogram that this way of selecting the sample results in large variation in the sampling weights. This large variation in the sampling weights will unduly inflate the variances of the estimates, it would have been better to use a design with less variation in the sample take within EA.

250



Figure 1: Sampling Weights for VHLSS 2002

# 3.5.3. Sampling Weights for VHLSS 2004

The sampling weights for the small sample used for the expenditure module in VHLSS 2004 (9,189 households) are shown in Figure 2. The group of very high weights at the right end of the histogram is the weights for urban EAs in province 701 (HCMC). The very large weights reflect the fact that the sampling fraction is very low in HCMC due to the square root allocation of the sample to provinces.



Figure 2: Sampling Weights for VHLSS 2004

The lowest group of weights is for rural EAs in provinces 301 and 606. The allocation of sample communes to these provinces is higher than what the square root allocation will give. Example: in province 301-rural, the square root allocation calls for 18 communes to be selected from this stratum but the actual sample was 30. The province 301 has been split into two provinces between 2002 and 2003; the two parts seem both to have kept the original sample size, resulting in heavy over-sampling especially in 301 rural.

# 3.5.4. Sampling Weights for the 2002-2004 Panel Data

About a half of households in VHLSS 2002 are resurveyed in VHLSS 2004. The question is how to use the weight for these households when using the panel data? Mr. Hans Petterson, who designed the sampling for VHLSS, suggests that it is better to use the weights of the VHLSS 2004 for the panel data as we have the most updated information about the population as well as a new sampling procedure for these households.

# 4. Implementation

# 4.1. The Central Level

The director of the Social and Environment Statistics Department of GSO has the overall responsibility for the implementation of the VHLSSs. This department is responsible for organizing, managing and monitoring all stages of the survey implementation, including data need assessment, questionnaires and manual design, sampling design, training of trainers from provincial statistics offices, monitoring of the training of interviewers, team leaders and field work. In addition, this department is also responsible for checking the quality of the data, data cleaning, analysis and dissemination.

## 4.2. The Provincial Level

The director of the provincial statistics office (PSO) is responsible for the organization and implementation of the survey within his province, including personnel arrangements, training of interviewers and team leaders, field work, monitoring of the implementation of the teams, data entry, and data checking and cleaning. In addition, he is responsible for the quality of the data. At the beginning of each round, it is recommended that the provincial statistics office sends the supervisors to the field in order to supervise the team about the problems that they meet in the field.

### 4.3. The District Level

The director of the statistics office in the district is responsible for implementing the survey in the selected commune within his district. Each district has one team of 3 to 4 people and a team leader, which is often the director or deputy director of the statistics office in the district.

### 4.4. The Commune/Ward Level

Leaders of the selected communes or wards are responsible for cooperating with the survey team, helping them to complete their work and ensure high data quality. The commune or ward leaders often organize the meeting with the selected households to inform about the purpose of the survey and the responsibility of the household according to the Statistics Law in order to get a good cooperation with the interviewer during the interview time. In the remote and mountainous areas, this work is done by village leaders. The commune/ward leaders and the team leader of the survey have set up a work plan for the field work in advance, so that they can arrange the meeting with the respondents.

#### 4.5. Recruitment of Interviewers and Team Leaders

This survey is considered as the most complicated survey in Vietnam so the quality of the interviewers and team leaders is a very important factor contributing to the quality and success of the survey. The GSO recruits these people based on education, experience, capacity in public relations, enthusiasm and health status.

### 4.6. Training of the Trainers

Two 10 day training courses for the trainers (one in the North and one in the South) are implemented for one leader and two staff members from the Social and Environment Statistics unit of PSO. The trainers come from the Social and Environment Statistics Department of GSO who are the designers of the questionnaires, manual and sampling for this survey. During the training time, there is one day for students to implement the pilot interview and an exam is given to the participants at the end of each training course to evaluate the quality of each potential trainer.

## 4.7. Training of the Interviewers and Team Leaders

The participants from the above mentioned training courses are the trainers for their province. Each province has the same training course with the same content as the training course for trainers. The participants of this training course are the interviewers, team leaders and supervisors.

#### 4.8. Media Campaign

Each selected household receives a letter from the GSO introducing the objectives of the survey and

the importance of the information that the household provides to the interviewer as well as the responsibility of the household in terms of provision of this information according to the Statistical Law. In addition, in each center of the commune or village (the cultural house), GSO posts a poster about the survey to increase the awareness of the public.

### 4.9. Field Work

The field work is done in about one month for each round of the survey. The VHLSS 2002 was implemented in 4 quarters. Each long questionnaire was completed within one-and-a-half day and the short questionnaire was completed within one day. In 2004, the survey was implemented in May and September. However, most of the first round was implemented in June and some provinces completed the field work for the second round in late 2004 (Ho Chi Minh City). The long questionnaire (including two special modules) was often completed within 2.5 days while the short was completed in one-and-a-half day. There were about 1,500 interviewers, 600 team leaders, 200 provincial supervisors and 20 central supervisors mobilized in 2002. The GSO also recruited a number of reserve staff for this survey to replace survey members that could not work for the survey due to health or other problems.

# 5. Data Processing

The data is first checked by the team leader before it is send to the PSO. The PSO checks the data again and subsequently send it to the data entry staff. The data entry staff is trained in using the data entry program (CsPro) by the Center of Information and Statistics. In the CsPro program, there is the module for data cleaning, which gives a warning for inconsistent or out of range input in each question. Based on these warnings, the PSO checks whether the errors are made by the computer operator or by the interviewer. For errors made by interviewers, the PSO sends them to the district and asks the interviewers to check the errors with the households and then return the revised input back to PSO.

After the data entry is completed at the PSO, it is sent to the Center of Information and Statistics as well as to the Social and Environment Statistics Department. The second data cleaning process is done at the Social and Environment Statistics Department. The main errors at this stage are the inconsistent error and out of range errors. There are about 5 or 6 people (each in charge of one or two sections) checking the data to find errors. In case of errors, the data are returned to the PSO for verification. Finally, it is sent back to the Social and Environment Statistics Department Statistics Department for the final checking. Due to the size of the survey, the total time for data cleaning is often from 8 to 12 months.

## 6. Dissemination Policy

# 6.1. Principles in VHLSS Data Dissemination

The policy of VHLSS data dissemination should follow a number of principles to ensure the legality of data dissemination, and the effectiveness of data use.

### 6.1.1. Legality of Data Dissemination

Dissemination of statistical information is one of the important functions of the GSO. As decreed in Article 1 of Degree 23/CP dated March 23, 1994 by the Government regarding the functions, duties, powers and organizational system of the GSO, the GSO is an agency under direct authority of the Government, with the function of State management in the field of statistics and provision of quantitative information on the socioeconomic situation to every agency, organization and individual

in accordance with the Government's regulations. The data and information collected from the VHLSS will be made available to the public in due time as regulated by present and future legal documents.

### 6.1.2. Equitability in Data Dissemination

In collecting, analyzing, and disseminating data, GSO adheres to the duties and powers of an agency of the Government that have evolved to protect the impartiality and credibility of federal statistical efforts. The GSO strives for equitable policies and practices on data dissemination, ensuring that the VHLSS results are revealed to all potential users including individuals and organizations both domestic and international. Everyone who meets the conditions specified in the data dissemination policy can obtain the data they need. The GSO tries the best to make the VHLSS available in an open environment, with full documentation of sampling method, questionnaire and data use introduction, summary results, and micro-data.

#### 6.1.3. Maintenance of Confidentiality

The Law on Statistics of Vietnam that was passed by the National Assembly and in effect since January 1, 2004 requires that confidential data not to be disseminated to the public. According to the Law, the confidential information includes: (i) Information on identification of individuals and organizations such as name and address. This identification information can be disseminated only in case of the respondents' agreement. (ii) Confidential information of the State.

Dissemination of the summary results from the VHLSS does not have a problem of ensuring confidentiality because the socioeconomic indicators are reported at the regional and provincial level. However, when revealing the micro-data, one should be cautious about the confidentiality of respondents. The VHLSS contains a wide range of information on living standards including sensitive personal information. The disclosure of sensitive information can have harmful impacts on respondent households, or even on communes. Although the identification information is deleted before micro-data dissemination, it is possible that the analysis of micro-data can locate groups of households, and especially communes. This is a central challenge to GSO to implement the dual mandates of maximizing the availability of micro-data while protecting confidentiality. Finding ways to make the VHLSS data available in sufficient detail for analytic purposes will often mean stretching the limits of data dissemination up to –but not beyond– the point where confidentiality is jeopardized.

### 6.1.4. Guarantee of Data Quality

According to the Law on Statistics the statistical information disseminated by the GSO has the highest legal effects. The GSO is fully responsible for objectivity and accuracy of the collected information. This puts a pressure on the GSO in ensuring the quality of the VHLSS because this is a large-scale survey with a large number of questions on living standards. Although the VHLSSs are implemented using extensive quality control procedures such as scientific questionnaire design, decentralized organization of fieldwork, enhanced interview skills, sound sampling method and data management, the evaluation of data quality prior dissemination is found very necessary. Analytical results of the VHLSS should be compared with results obtained from other surveys conducted by the GSO and information sources of different ministries. Preliminary results of the VHLSS should also be presented to various ministries and international cooperation agencies for comments on the reliability of the data before the completed results are released to the public.

# 6.2. The Policy of VHLSS Dissemination

Data and information collected from VHLSS are very useful for the monitoring and evaluation of living standards of people, assessment and design of socioeconomic policies and programs. As a State

agency in the statistical system, the GSO has the obligation to maximize availability of the VHLSSs to the public, minimizing the time from data collection to dissemination so as to maximize the usefulness of the VHLSS. At the same time the data dissemination has to ensure the confidentiality and follow the Government's regulations. Final decisions about VHLSS dissemination will depend on the specific data collected in each survey, and can only be made in detail after the data have been collected, processed, and reviewed for the confidentiality and legal framework of data dissemination issues. The policy of data dissemination proposed just provides general guidelines for the dissemination of the VHLSS.

### 6.2.1. Data to Be Disseminated

The GSO aims to make the data and information from VHLSS available in user-friendly formats for all sorts of users, subject only to limits imposed by data quality, legal regulations, and the need to protect confidentiality. The following documents and data will be disseminated for users depending on their need.

1. Basic documents of VHLSS: They include the questionnaires, format files, data dictionaries, code books, brief descriptions of any constructed variables made available and brief sampling information. For some users, more detail will be required, such as the manuals for the interviewers, supervisors or data entry operators, and other information related to the survey design and data collection. These documents are prepared in printout and electronic formats (CD-ROM or diskette) depending on the users' requirements.

2. Preliminary reports of VHLSS results: Information collected from VHLSS are analyzed to produce statistical abstracts and tabulation tables on general socioeconomic indicators: poverty and inequality, demography, income and consumption expenditure, education, employment, health, housing and durable goods, participation in poverty alleviation programs. These welfare indicators are estimated for the whole country, separately for urban and rural areas, and for eight socioeconomic regions. The regions consist of: the Red River Delta, the North East, the North West, the North Central Coast, the South Central Coast, the Central Highlands, the North East South, and the Mekong River Delta. The reports have the length of 30 to 50 pages, and are made in printout and electronic formats (CD-ROM or diskette) depending on the users' requirements.

3. Completed reports of VHLSS results: These reports provide detailed information on socioeconomic characteristics of households and communities. In addition to general indicators of socioeconomic issues as given by the preliminary reports, the completed reports present the detailed analysis of each issue. For example, while the preliminary reports provide only estimates of the income level, the completed reports break down income by income sources, household size, gender and age of household head. The results are estimated at the regional and provincial level. The reports have a length from 80 to 120 pages, and made in printout and electronic formats (CD-ROM or diskette) depending on the users' requirement.

4. A book of VHLSS results: After receiving comments on the completed reports, the GSO will revise and add more detailed results of statistical analysis of living standards using VHLSS data to produce a book "Results of Viet Nam Household Living Standards Surveys". The book contains from 10 to 15 chapters with a length of around 500 pages, providing insight into the monitoring and evaluation of living standards, trends in poverty and inequality using data tables, illustrative figures and charts. It is printed in both Vietnamese and English, and published by the Statistical Publishing House.

5. Required statistical products: Being aware that the micro-data is difficult to get used to, the GSO will provide results of the statistical analysis as required by particular users. The users can request detailed analytical reports based on VHLSS data including cross-tabulation tables, figures,

charts, and statistical interpretation. The required reports are made in the form, printout or electronic form (CD-ROM or diskette), depending on the users' needs.

In order to increase the effectiveness of the data use the GSO will conduct presentations and training courses related to the use of the VHLSS as required by the users. Technical assistance in analyzing the VHLSS data is also provided to help the users reduce their time in getting used to the data and increase the efficiency and science of their research.

6. Micro-data set: After deleting the identification variables to protect the confidentiality, the micro-data set of VHLSS including the household and commune data will be provided for the users. In general, the data are grouped into different files by the thematic sections in the questionnaire, e.g. education or employment. Identity codes for respondents including household members, households, and communes are generated to facilitate merging of the data between data files. The codes can be used to link the data from different VHLSS to create the panel data. The whole data set or a portion of it can be provided for the users depending on their need. The GSO also prepares aggregated files with variables selected by the users.

The micro-data of VHLSS are stored using the program "Stata". The Stata files can be easily transformed into other file formats such as SPSS, EXCEL, SAS, etc using data-transfer programs.

#### 6.2.2. Dissemination Calendar

The GSO makes every effort to disseminate the survey documents, reports and micro-data of VHLSS as soon as possible following data collection, subject only to limits imposed by resources, technology, and data quality. The GSO will not restrict the dissemination in order to preserve publication rights of its staff, donor agencies, or other State organizations. Prior to the dissemination of the VHLSS the GSO will thoroughly evaluate data quality and assure that the data release will preserve the respondents' confidentiality. Expert assistance is often needed to conduct data quality reviews. Procedures for conducting such reviews should be a part of all data planning activities.

1. For basic documents of VHLSS: The basic documents on the survey design and data collection such as questionnaire, sampling methods, interview manuals will be provided for the users after the data collection is completed. Other documents related to the data processing and cleaning are disseminated as soon as they are completed.

2. For result reports: VHLSS is a large survey collecting a huge amount of information. It is estimated that the preliminarily processing and cleaning of data might take from four to six months. The statistical reports using VHLSS data on topics required by particular users are expected to be provided for the users four months after the completion of data collection. The preliminary reports will be disseminated six months after the completion of data collection. Then preliminary reports will be revised and added more detailed information on living standards to produce completed reports. These reports are released nine months after the data collection is completed.

3. The publication of the book "Results of Viet Nam Household Living Standards Surveys" requires more time for data analysis, editing and printing. The book is planned to be published one and half year after the completion of data collection.

4. For micro-data: The GSO strives for making the VHLSS micro-data set available as promptly as possible, subject only to limits imposed by resources, technology, and data quality. As a State agency, the GSO has to be fully responsible for the disseminated data. They are required to ensure the quality of VHLSS data and protection of respondent. For the living standard survey, the requirement of micro-data equality is utmost because the analysis of micro-data results in socioeconomic implications which might have strong effects on the living standards of the people. The VHLSS micro-data is reviewed carefully by the GSO and assistance agencies by comparing the results extracted from these data with other information sources. When there are suspicions about the accuracy of data, the GSO will have to review the data from the collection to the computer entry. If the data cannot be

corrected they have to be deleted, and the weight might be adjusted to ensure the representativeness of the data set. The evaluation of data quality is conducted during the processing and analyzing of the data. This substantial work can take a long time. The micro-data of VHLSS are proposed to be released to the users two years after the data collection is completed.

There are, however, situations where it would be beneficial to release a portion of the micro-data or aggregated data prior to the time when the full set of micro-data can be made available. The requirement for such "early releases" or a staggered release can be raised by the GSO and donor agencies to fulfill important policy and scientific goals.

#### 6.2.3. Recipients and Dissemination Mechanism

The GSO aims to disseminate the data and information of VHLSS as widely as possible, subject only to limits imposed by resources, technology, data quality, and confidentiality protection. All kinds of people are ensured equitable accessibility to the reports and micro-data of VHLSS.

For the document and analytical reports of VHLSS: The survey documents and reports of VHLSS results including preliminarily and completed reports as specified above are made available to the public. However in the context of limited human and material resources the GSO will require the users to meet provision conditions to assure that the users do not misuse the data and not ask for more data than they are likely to use. The conditions are applied as follows:

• For State organizations, international agencies with offices in Vietnam, domestic institutes and universities and their staffs, researchers and students, the documents and reports are provided if they hand in a formal letter of data request. This letter should State clearly the purpose of data use. A small fee will be charged to cover the costs of printing and mailing.

• For other users, in addition to the fee of printing and mailing a modest fee of data distribution will be charged depending on the required amount of documents and reports in order to prevent misuse of data that could affect the human and material resources of the GSO. The fee does not include the cost of data collection, computer entry, data processing and cleaning, and data analysis.

The book "Results of Viet Nam Household Living Standards Surveys" is published for sale.

Required statistical products: The GSO provides particular analytical reports, statistical assistance, presentations and training courses according to users' requirements. A service fee to cover the cost of data analysis, report writing, and assistance will be charged to the users. The amount of fee payment will depend on the specific requirements of the users.

Micro-data set: The micro-data set of VHLSS will be disseminated to the users on the basis of user payments with two access policies: data use agreements and a controlled access to micro-data:

• Data Use Agreements: The micro-data are only provided for the users if they commit themselves to data use agreements. Data which cannot be publicly released will not be made available without the data user's signed written agreement to provide necessary safeguards. In general the data use agreements include the following: (i) The users are prohibited from modifying and changing the micro-data. (ii) Research reports using the VHLSS data are required to specify the data source from VHLSS. (iii) The users are not allowed to use the micro-data to track the identification of the respondents. (iv) The users are required not to pass the data set to third parties for any reason. (v) If the users obtain the data set without payment, they have to hand in a request for using the data, and commit themselves to using the data only for the specified purposes. Further use of the micro-data for new researches will need the approval of the GSO. A copy of research reports using the VHLSS data is required to be sent to the GSO.

• Controlled access: If the users conduct research assignments for the GSO and donor agencies that rely on the VHLSS micro-data, users will have access to the data without any payment but are not in possession of the data. The GSO and donor agencies exercise direct supervision of the data use in

order to protect the misuse of data. When the assignments are completed the GSO might ask the users to return the data set.

For the State organizations, the micro data will be provided through the mechanism of data use agreements two years after the data collection. A small fee will be charged to cover the cost of producing CD/diskette and delivery.

For other users, the full micro-data set will be supplied on a payment basis. The users are required not to track the respondents' identification, to modify the micro-data, and to pass the data to third parties. The data processing and cleaning require substantial resources. Thus the charge aims to cover a fraction of costs of sorting and cleaning information, and distributing the data set. In addition, the charge also helps to ensure that the use of the micro-data is necessary and the potential users have an incentive not to ask for more data than they are likely to use. The cost of data collection is not included in the charges.

**Publishing Responsibility** BUI VIET BAC

# **Tener Responsibility** PHAM NGOC LUAT

Editor NGUYEN PHUONG MAI

**Cover** DUONG HONG MAI

Printed 800 copies, size 21x28 cm, in National Political Publisher. Publishing plan: 871-2007/CXB/03-149/VHTT. Publishing decision: 2373-QĐ/VHTT, dated December 31st 2007. Completed and Registered in February 2008.