Inequality, Rural Development and Food Policy: Essays in Applied Economics and Microeconometrics

by

Hoa Thi Minh Nguyen

A thesis submitted for the degree of
Doctor of Philosophy

at the
The Australian National University

October 2012

Declaration of Authorship

I, Hoa Thi Minh Nguyen, declare that this thesis entitled, 'Inequality, Rural Development and Food Policy: Essays in Applied Economics and Microeconometrics' and the work presented in it are my own, except where otherwise indicated. I confirm that:

- Parts of Chapter 2 were published with my thesis supervisors as Breusch, T., M. Ward, H.T.M. Nguyen, and T. Kompas (2011), 'On the Fixed-Effect Vector Decomposition', *Political Analysis* 19(2), 123–134. In this research, I initially discovered the potential errors with the Fixed Effects Vector Decomposition when I worked on ethnic inequality for Chapter 1 of my thesis. I provided all of the empirical and Monte Carlo evidence for the theorems, and made substantive contributions to all sections of the work, except for the Monte Carlo evidence comparing FEVD with alternative estimators.
- Chapter 3 has been published as Kompas, T., N.T. Che, H.T.M. Nguyen, and H.Q. Nguyen (2012), 'Productivity, Net Returns and Efficiency: Land and Market Reform in Vietnamese Rice Production', *Land Economics* 88(3), 478–495. In this work, I conducted the data analysis and estimations for all of the stochastic frontier and inefficiency models. I also provided the literature review, updated the supporting data set and assisted with the calculations for the analysis of total factor productivity, terms of trade and net returns. In terms of exposition, I constructed the sections on context and stochastic frontier and inefficiency models, and made substantive contributions to all other sections of the paper. H.Q. Nguyen provided the farm survey data.
- In Chapter 4, Pham Van Ha provided output from the Computational General Equilibrium (CGE) model and T. Bui provided the input-output (data) tables used for the CGE modelling.

Signed: Hoa Nguyễn Thị Minh Hoa

Date: 10/10/2012

Acknowledgements

First and foremost, I offer my sincerest gratitude to my principal supervisor, Professor Tom Kompas. His genius, his encouragement and his guidance opened research doors and ways of learning for me that I never would have thought possible. I am also deeply indebted to my co-supervisor, Professor Trevor Breusch, for his support, advice and unsurpassed knowledge of econometrics, not to mention his endless patience with my work. I would also like to thank Professor Michael Ward, my third panel member, now at Monash University, for his suggestions and assistance on various estimation methods, all of which were essential components to much of my econometric analysis. I am truly proud to have a 'stamp of approval' on this work from all three of these supervisors, and for the opportunity to work so closely with them throughout my time at ANU.

I am very grateful for research and career advice along the way from Dr. Martin Rama, Professor Henrik Hansen and Professor Quentin Grafton. Comments on my work received from Professor Bruce Chapman, Professor Stephen Howes, Dr. Amy Liu, Doan Thi Thuy Dung, La Hai Anh, Dr. Brian McCaig, and participants at the Crawford School PhD Seminar in Economics, the Crawford School PhD Annual Conference, the 2nd Vietnam Economic Workshop at the ANU, the 8th Australasian Development Economics Workshop at Monash University, and the Econometric Society Australasian Meetings in Melbourne (2012) are all gratefully acknowledged.

My research over the years benefited from helpful discussions with my colleagues at the World Bank Office in Hanoi, especially Dr. Nguyen Nguyet Nga, Dr. Doan Hong Quang, Carrie Turk, Ngo Thi Quynh Hoa, Nguyen Thi Thu Lan, Rob Swinkels and Dr. Vivek Suri. I also sincerely thank my colleagues at the Centre for Analysis and Forecasting, Vietnam Academy of Social Sciences, especially Dr. Nguyen Thang, Dr. Nguyen Thi Thanh Ha, Cao Thi Thuy, Tran Ngo Thi Minh Tam, Nguyen Thi Thu Phuong, Le Dang Trung, Nguyen Thi Van Ha and Hoang Thanh Tu for their helpful comments and assistance.

My study in Australia was very nicely enhanced by the wonderful professional and academic staff at the Crawford School of Public Policy. I want to especially thank Jenny Colman for her kind assistance throughout my study, Robyn Walter for her administrative guidance, Dr. Wendy Noble for her academic skills advice and

Tony Varty and Jay Giampieri for their IT support. I also thank Krysthea Palade, Alison Cumming-Thom, Janine O'Flynn, Billie Headon, Sue Farrow, Shuqun Zhao and Ngan Le, all of whom have been very kind and helpful in their respective roles. I would also like to express my sincere thanks to my best friends, Dr. Chu Hoang Long, Dr. Phung Duc Tung and his wife Nguyen Thi Viet Nga, Nguyen Thi Thu Thuy, Pham Huong Giang, Dr. Le Quang Thanh, Dang Duc Anh, Nguyen Thi Minh Hien and Dr. Nguyen Tam Giang for their enormous support and wonderful friendship. Thanks also are expressed to Megan Poore and my fellow PhD students, Undargaa Sandagsuren, Marija Taflaga, Yu Wei, Sandra Velarde and Mary Milne for sharing experiences and many happy lunch times in the PhD common room.

My study at ANU was made possible by generous scholarships and financial support provided by the Government of Australia, the Crawford School of Public Policy and the Commonwealth Environment Research Facilities. I greatly appreciate this support.

Last but not least, this thesis would have remained but a dream had it not been for the immense patience and confidence provided to me by my relatives: my dearest daughter Vu Thao Ngoc, my late father Nguyen Huu Chung, my mother Nguyen Thi Minh Duc, my uncle Nguyen Khanh Hoa and his wife Trang Thi Dai, my sisters Nguyen Thi Minh Chau and Nguyen Thi Minh Hien, my late brother Nguyen Truong Giang, my brother-in-law Pham Viet Hung, my nieces Nguyen Thi Cam Nhung, Pham Minh Anh, Pham Viet An and my nephew, Nguyen Trung Hieu. To all of you I dedicate this work.

Abstract

The thesis contains four essays on 'Inequality, Rural Development and Food Policy in Vietnam'. Re-examining the sources of ethnic inequality, the first essay uses instrumental variable approaches to provide consistent estimators of explanatory variables at household and commune levels for ethnic differences in household expenditure per person. Four key conclusions are drawn. First, removing language barriers significantly reduces ethnic inequality, especially through enhancing the gains earned by minorities from education. Second, variations in returns to education favour the majority in mixed communes, suggesting that special needs of minority students have not been adequately addressed, or that there exists unequal treatment in the labour market. Third, with the exception of hard-surfaced roads, there is little difference in the benefits drawn from enhanced infrastructure at the commune level across ethnic groups. Finally, contrary to established views, we find that as much as 49 to 66 percent of the ethnic gap is attributed to differences in endowments, not to differences in the returns to endowments.

The second essay analyses the properties of the fixed-effects vector decomposition estimator, an emerging and popular technique for estimating time-invariant variables in panel data models with group effects. The formal analysis finds: (1) This decomposition is equivalent to a standard instrumental variables approach, for a specific set of instruments. (2) The estimator reproduces classical fixed-effects estimates for time-varying variables exactly. (3) The standard errors recommended for this estimator are too small for both time-varying and time-invariant variables. (4) The estimator is inconsistent when the time-invariant variables are endogenous. (5) The reported sampling properties in the original Monte Carlo evidence do not account for the presence of group effects. (6) The decomposition estimator has a higher risk than existing shrinkage approaches, unless the endogeneity problem is known to be small or no relevant instruments exist.

The third essay examines the effects of extensive land and market reform in Vietnam on rice output and incomes, principally illustrated with measures of total factor productivity, net incomes and net returns in rice production from 1985-2006. Results also show considerable gains in major rice growing areas, but recent evidence of a productivity 'slow down'. The differences over time and region speak to existing land use practice, calling for further reform. Stochastic frontier estimations detail the effects of remaining institutional and policy constraints, including existing restrictions on land use, ambiguous property rights, and inadequate markets for land and access to extension services and credit.

The fourth essay analyses Vietnam's rice export policy and recent export ban in the context of rising food prices, drawing on insights from a regionally-disaggregated or 'bottom-up' CGE model and a micro-simulation using household data. Three main conclusions are drawn. First, although there is little impact on GDP, there are substantial distributional impacts across regions and households from different export policies and market conditions. Second, both rural and urban households, including poor households, benefit from free trade, even though domestic rice prices are higher. Finally, under free trade, relatively large gains accrue to rural households, where poverty is most pervasive.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Vietnam's performance over the last three or four decades has been regarded as one of the most successful stories in economic development of the last century. Since 1986, when Vietnam made its landmark commitment, under *Doi Moi*, to transform the economy from a centrally–planned model into a market-based economy with a socialist orientation, it has achieved remarkable progress in economic growth and poverty reduction. Gross Domestic Product (GDP) per capita in 2008, for example, was more than three times larger in comparison with that at the outset of *Doi Moi* (General Statistic Office, 2000, 2009), and between 1993, when the first household expenditure survey was conducted, and 2008, the poverty rate among the population as a whole fell sharply from 58 to 14.5 percent (Vietnam Academy of Social Sciences, 2011), lifting more than 30 million people out of poverty.

Much of the remarkable and ongoing progress in reducing rural poverty, in particular, has been achieved largely through a series of extensive market and land reforms in agriculture and rural institutions. The land and market reforms in agriculture were pervasive, moving the system of rice production from commune-based public ownership and control to one with effective private property rights over land and farm assets, competitive domestic markets and individual decision making over a wide range of agricultural activities. The substantial incentive effects created by these policy measures, inducing farmers to work harder and use land more efficiently, have been estimated to be as much as fifty percent of the increase in total factor productivity (TFP) during the peak of the reform period

(Che et al., 2006). Overall, given these reforms, Vietnam has gone from being a large importer of rice in 1976-80, to now the second largest exporter of rice in the world, with considerable increases in farm profitability and rural incomes, and decreases in rural poverty rates by over forty percent from 1993 to 2004 alone (Hansen and Nguyen, 2007).

However, much still remains to be done to increase living standards in rural areas and enhance general rural development. Like many reform processes, the early rapid gains in economic activity have dissipated over time, with the suggestion now of a TFP 'slow down' in rice production in many areas in Vietnam. In addition, many of the poor still farm small areas of land, constrained in use, often with fragmented or non-contiguous plots, and with little or no human and physical capital accumulation or access to agricultural extension services and farm credit. Much of this is due to remnants from past institutional arrangements, but also to continued constraints in land use, credit availability and the provision of rural services, all calling for further or renewed land policy and market reform.

Along with dramatically changing institutions and economic conditions, with *Doi Moi*, as well as international and regional integration, inequality has also increased in Vietnam, although at a comparatively modest rate by international standards. Gaps in income, especially between rural and urban areas, are increasing while remaining poverty is characterised by strong spatial or regional dimensions. For example, in 2008, the poverty rate in the South East of Vietnam was 3.5 percent while in the North West it was 45.7 percent. At the same time, household per capita expenditure and income of an average urban household is almost double those of counterparts in rural areas. This fact, together with disparities in other socio-economic aspects, has motivated many people to migrate from rural to urban areas in spite of constraints on movement and other administrative barriers. This migration can only be slowed by added and significant economic development in rural areas.

Further reductions in rural poverty in the years to come will likely still depend on policies in the rice sector. While contributing less than 20 percent to GDP, this sector is of particular political, social and economic importance, and for good reason. Vietnam is the world's second largest rice exporter, with exports of six million tons, equivalent to about 16 percent of the world trade volume in rice (Shigetomi et al., 2011). While export revenues represent only roughly three percent of Vietnam's GDP (World Bank, 2009c), production and sales in this sector have a broad distributional impact since as much as 66 percent of rural households and 77 percent of the poorest quintile in Vietnam are rice producers (General Statistic Office, 2006a). Rice is also the dominant staple food of the Vietnamese, and accounts for about 33 percent of the household expenditure among the poorest quintile households (General Statistic Office, 2006a).

Given the importance of rice, the Government of Vietnam has maintained strict control over rice exports since 1992, three years after Vietnam began exporting rice. To do so, the Government sets annual rice export targets, which can be adjusted throughout the year, subject to changes in domestic supply and demand. The Government, at its discretion, can also suspend rice exports whenever it is deemed necessary. This control is further underpinned by the near 'monopoly power' possessed by state-owned enterprises (SOEs) in the domestic rice export market, as well as their heavy involvement in rice export policy formulation and management. SOEs also dominate north-south trade in the domestic market, especially under the national food security framework, in spite of representing only 25-30 percent of the total domestic rice market. Given SOE control, generally partitioned by north and south, Vietnam's domestic rice supplies are characterised by a lack of integration between markets across the north and the south of the country (Minot and Goletti, 1998; Baulch et al., 2008; Luu, 2003).

Vietnam's international integration, especially the accession to the World Trade Organisation (WTO) in early 2007, has brought many opportunities as well as challenges for the rural Vietnamese. One of the key risks that farmers face is price uncertainty in world prices for commodities. At times, the fluctuations in world commodity prices can be dramatic, resulting in large swings in farm and household income. For example, world food prices reached their peak in the second quarter of 2008, with wheat and maize three times and rice five times more expensive than at the beginning of 2003 (Von Braun, 2008). In the face of rising world prices for rice, and given its importance in Vietnam, the government imposed an export ban from March 25 to the end of June. By the time the ban was lifted, the world price for rice had begun trending downward. This downward trend, coupled with a domestic excess supply, dramatically pushed rice farmers' profits down in what would have otherwise been a 'golden year' for rice production and profitability.

The rice export ban by the government reveals shortcomings in existing governance and institutions. While the objective of ensuring domestic food security and controlling inflation was appealing, poor price forecasts, lack of transparency and market distortions generated considerable uncertainty and less than optimal behaviour, preventing rural households, many of them poor, from making full use of international integration.

In terms of the progress toward poverty reduction in Vietnam, challenges especially remain for ethnic minorities. Although positive developments are broadly visible across different segments of the population, the gains from growth have not been shared proportionately among different groups of people. For example, while the poverty rate of the Kinh and Chinese fell from 54 percent in 1993 to 9 percent in 2008, for other ethnic minorities as a whole, it decreased more modestly, from 86 to 50 percent over the same period of time. The proportion of ethnic minorities among poor households has also increased sharply from 17.7 percent in 1993 to 40.7 percent in 2008, despite representing only 14 percent of the country's population (Vietnam Academy of Social Sciences, 2011). Together, ethnic minorities account for more than 60 percent of those classified as 'hungry' in Vietnam.

The Government of Vietnam has a number of policies and programs in place to help the minority group. These policies and programs are based on two approaches, those that target communes and those that target households. As an example of the former, Program 135 largely finances local infrastructure improvement (e.g., the provision of roads, power and water) in so-called "extremely difficult" communes in ethnic minority and mountainous areas. For the latter, the Hunger Eradication and Poverty Reduction Program, targets poor households (largely the minority), by providing access to credit, exemption from education fees, and support for health care, among other benefits. In spite of these policies and programs, progress in raising the living standards of the minority has been much slower than that for the majority.

Inclusive development is at root of social and political stability, which the Communist Party of Vietnam has vowed to protect. Therefore, inclusive development must remain at the heart of Vietnam's development strategies. Looking forward, whether or not this objective is realised, may well depend on Vietnam's performance in rural development and poverty reduction among ethnic minorities in the years to come.

1.2 Outline of the Thesis and Key Results

Against the background, this thesis focuses on two broad issues, (a) ethnic inequality and (b) rural development and food policy with a particular interest in the rice production and export sectors. These two broad issues are discussed in four stand-alone essays in applied economics and microeconometrics.

The first essay examines what drives the gap in the living standard between the Kinh and Chinese (defined as the 'majority' in this thesis) and other ethnic minorities as a whole (defined as the 'minority'), measured by differences in household expenditures per person. In particular, we investigate the role of language barriers and how they may hinder minority households from taking advantage of their acquired skills and attributes; whether commune infrastructure, a key instrument used by the Vietnamese Government to narrow the ethnic gap, works for or against the minority; and to what extent preferential treatment and differences in endowments, as opposed to returns to endowments, explain the ethnic gap in expenditures.

Using data from a household survey in Vietnam in 2006 (VHLSS 2006), we draw four key conclusions. First, removing language barriers would significantly reduce inequality among ethnic groups, narrowing the ethnic gap, and especially so through enhancing the gains earned by minorities from education. Second, variations in returns to education exist in favour of the majority in mixed communes, suggesting that either the special needs of minority children have not been adequately addressed in the classroom, or that there exists unequal or preferential treatment in the labour market. Third, in contrast to recent literature, there is little difference between ethnic groups in terms of the benefits drawn from enhanced infrastructure, such as power and clean water, at the commune level. An exception is the returns to paved or hard-surfaced roads, which differentially benefits the minority group. Finally, contrary to established views, we find that as much as 49 to 66 percent of the ethnic gap is attributed to differences in endowments, and not to differences in the returns to endowments.

The second essay analyses the properties of a recently introduced methodology for panel data, known as fixed-effects vector decomposition or FEVD, which Plümper and Troeger (2007a) developed to produce improved estimates for such time-invariant or slowly-changing variables. I came across this methodology in a search for an approach which could provide consistent estimators of explanatory variables at household and commune levels for ethnic differences in household expenditure per person in the first essay of this thesis. As households in VHLSS 2006 are sampled by communes, they share common commune-level effects, thereby forming a panel-like data structure.

Researchers in many fields seek to exploit the advantages of such panel data. Having repeated observations across time for each group in a panel allows one, under suitable assumptions, to control for unobserved heterogeneity across the groups which might otherwise bias the estimates. Mundlak (1978) demonstrated that a generalized least squares approach to unobserved group effects, which treats them as random and potentially correlated with the regressor, gives rise to the traditional fixed-effects (FE) estimator. However, FE is a blunt instrument for controlling for correlation between observed and unobserved characteristics because it ignores any systematic average differences between groups. Thus any potential explanatory factors that are constant longitudinally (time-invariant) will be ignored by the FE estimator. Likewise, any explanatory variables that have little within variation (that is, slowly-changing over the longitudinal dimension) will have little explanatory power, and will result in imprecise coefficient estimates that have large standard errors.

Hausman and Taylor (1981) had previously shown that a better estimator than FE is available if some of the explanatory variables are known to be uncorrelated with the unobserved group effect, thus described as *exogenous* explanatory variables. The Hausman-Taylor (HT) estimator is an instrumental variables (IV) procedure that combines aspects of both fixed-effects and random-effects estimation. Given a sufficient number of exogenous regressors, the HT procedure allows time-invariant variables to be kept in the model. It also provides more efficient estimates than FE for the coefficients of the exogenous time-varying variables. The downside of the HT estimator resides in specifying the exogeneity status for each of the time-varying and time-invariant variables in the model. In many practical applications such detailed specification is onerous.

Plümper and Troeger introduced FEVD as an alternative that seemed to be superior to HT because it requires fewer explicit assumptions yet seemed to always have more desirable sampling properties. Plümper and Troeger motivated the FEVD procedure on heuristic grounds, and advocated it on the strength of favorable results in a Monte Carlo simulation study. In particular, the simulation indicated that FEVD has superior sampling properties for time-invariant explanatory variables.

This second essay, in this sense, is a remedy to the lack of formal analysis. There are six main results. First, the FEVD is equivalent to a standard instrumental variables approach, for a specific set of instruments. In this sense, it provides nothing that is new or different to estimation techniques in this area. Second, the estimator reproduces classical fixed-effects estimates for time-varying variables exactly. Third, the standard errors recommended for this estimator are too small for both time-varying and time-invariant variables. Fourth, the estimator is inconsistent when the time-invariant variables are endogenous. Fifth, the reported sampling properties in the original Monte Carlo evidence do not account for the presence of group effects. Finally, the decomposition estimator has a higher risk than existing shrinkage approaches, unless the endogeneity problem is known to be small or no relevant instruments exist.

The third essay focuses on productivity, net returns and efficiency in Vietnamese rice production. This essay has two basic tasks. First, it assembles a data set from 1985 to 2006 to measure the changes in total factor productivity (TFP), terms of trade, net incomes and net returns in Vietnamese rice production, both in the principal rice growing areas and throughout the country. The results track the effects of the major land and market reform process and determine key differences in TFP and net returns over time and by region. All of this speaks directly to existing land use practices and is suggestive of needed policy response. The second task is to isolate the remaining institutional constraints and policy challenges that may be limiting increases in productivity and efficiency. For this purpose, two stochastic production frontier and inefficiency models, drawn from two different farm survey data sets, are estimated to determine the potential effects of ongoing issues over land use and sale, the provision of credit, land fragmentation, less than secure property rights and the lack of rural education and support services.

The work finds that extensive land and market reform in Vietnam has resulted in dramatic increases in rice output and incomes. This is illustrated with measures of TFP, net incomes and net returns in rice production from 1985-2006. Results show considerable gains in major rice growing areas, but recent evidence of a productivity 'slow down'. The differences over time and region speak to existing land use practice, calling for further reform. Estimations detail the effects of remaining institutional and policy constraints, including existing restrictions on land use, ambiguous property rights, and inadequate markets for land and access to extension services and credit.

In the forth essay, Vietnam's rice export policy is analysed in the context of rising world rice prices. In particular, the essay investigates both national and sub-national (or regional) impacts as well as distributional implications of different policy scenarios. To do so, insights from a regionally-disaggregated or 'bottom-up' computable general equilibrium (CGE) and a micro–simulation on household data are brought together . At the economy-wide level, the bottom-up CGE model is a combination of eight interacting CGE models, representing eight regions in Vietnam. To this end, it allows both national and subnational equilibrium assessments of an international rice price change to GDP, domestic prices and employment under different policy scenarios. Subnational changes in domestic producer and consumer prices of both food and non-food as well as factor prices generated from the bottom-up CGE are then inputed into household data for further disaggregated analysis of different policy options at the household level.

Three policy scenarios are considered in this essay. The first is when Vietnam maintains the status-quo with a rice export control, designed to mimic the imposed export ban in 2008, and the prevailing and market-segmenting powers of the SOEs in domestic rice markets. In the second scenario, Vietnam still controls rice export quantities, but liberalises the rice export market domestically — a WTO accession commitment Vietnam has promised to deliver since 2011. In the last scenario,we assume that Vietnam has a completely free rice export policy, with no export controls or bans, and a competitive domestic rice market.

Three main conclusions are drawn. First, although there is little impact on GDP, there are substantial distributional impacts across regions and households from different export policies and market conditions. Second, both rural and urban households, including poor households, benefit from free trade, even though domestic rice prices are higher. For many, this is a surprising result until it is recognised that many households produce rice in Vietnam, and wages often correlate with the price of rice. Finally, under free trade, relatively large gains accrue to rural households, where poverty is most pervasive in Vietnam.

Chapter 2

Language, Mixed Communes and Infrastructure: Sources of Inequality and Ethnic Minorities in Vietnam

2.1 Introduction

Inequality in wealth and income is often the source of tension between large disenfranchised groups of relatively poor minorities and the majority population. Failure to address this inequality may lead to ethnic conflict, resulting in poor economic performance and political instability (Easterly and Levine, 1997). Although ethnic inequality is not a characteristic of transitional economies alone, such concerns tend to predominate in these countries due to high but unequally shared growth in incomes, substantial differences in initial endowments and dramatically changing institutions and economic conditions that often quickly leave the poor behind.

Vietnam offers a useful case study in this regard. In the transition to a market-based economy, Vietnam has experienced remarkable success in economic growth and poverty reduction. GDP per capita in 2008 was three times larger than that in 1986, when Vietnam first made a landmark commitment to economic reform

(General Statistic Office, 2000, 2009). Between 1993, when the first household expenditure survey was conducted, and 2006, the poverty rate among the population as a whole fell from 58 to 16 percent.

Nonetheless, the gains from growth have not been shared proportionately among different groups of people. For example, while the poverty rate of the Kinh and Chinese (defined as the 'majority' in this paper) fell from 54 percent in 1993 to 10 percent in 2006, for other ethnic minorities as a whole (defined as the 'minority'), it decreased more modestly, from 86 to 52 percent over the same period of time (World Bank, 2007). Moreover, in 2006, the minority group accounted for 44 percent of the poor and 59 percent of those classified as 'hungry' in Vietnam, despite representing only 14 percent of the country's population (World Bank, 2007). The gap in expenditure between the two groups has also widened over time (Baulch et al., 2012, 2010).

The Government of Vietnam has a number of policies and programs in place to help the minority group. These policies and programs are based on two approaches, those that target communes and those that target households. As an example of the former, Program 135 largely finances local infrastructure improvement (e.g., the provision of roads, power and water) in so-called "extremely difficult" communes in ethnic minority and mountainous areas. For the latter, the Hunger Eradication and Poverty Reduction Program, targets poor households (largely the minority), by providing access to credit, exemption from education fees, and support for health care, among other benefits. In spite of these policies and programs, progress in raising the living standard of the minority has been much slower than that for the majority.

This paper examines what drives the gap in the living standard between the majority and minority groups, measured by differences in household expenditures per person. In particular, we investigate the role of language barriers[1] and how they may hinder minority households from taking advantage of their acquired skills and attributes; whether commune infrastructure, a key instrument used by the Vietnamese Government to narrow the ethnic gap, works for or against the minority; and to what extent preferential treatment and differences in endowments, as opposed to returns to endowments, explain the ethnic gap in expenditures.

[1]The language barrier here refers to the inability to speak Vietnamese. According to *Ethnologue* (as quoted in World Bank (2009a)), Vietnam encompasses seven major language families but 102 distinct languages. Vietnamese, the language of the Kinh, is the majority and official language.

Using data from a household survey in Vietnam in 2006, we draw four key conclusions. First, removing language barriers would significantly reduce inequality among ethnic groups, narrowing the ethnic gap, and especially so through enhancing the gains earned by minorities from education. Second, variations in returns to education exist in favour of the majority in mixed communes, suggesting that either the special needs of minority children have not been adequately addressed in the classroom, or that there exists unequal or preferential treatment in the labour market. Third, in contrast to recent literature, there is little difference between ethnic groups in terms of the benefits drawn from enhanced infrastructure, such as power and clean water, at the commune level. An exception is the returns to paved or hard-surfaced roads, which differentially benefits the minority group. Finally, contrary to established views, we find that as much as 49 to 66 percent of the ethnic gap is attributed to differences in endowments, and not to differences in the returns to endowments.

Our results are important for a number of reasons. First, they point to language as a significant determinant of the ethnic gap in expenditures. Traditionally, language has yet to be explicitly controlled for in econometric models examining ethnic inequality in Vietnam.[2] One reason for this could be the small sample size of household data with language variables in the first and early national surveys conducted in Vietnam.[3] But perhaps a more fundamental reason is that most studies use an Oaxaca-Blinder decomposition framework. Here, the differences in average outcomes in the two groups are decomposed into differences in the average level of each characteristic and differences in the returns to these characteristics between groups. Most of the researchers who use this method focus on variables that are relevant to both groups. As language barriers are almost entirely restricted to the minority, this variable is often dropped from the analysis.

In qualitative analyses, on the other hand, this is never the case. Language barriers are seen as key and have been highlighted as a major constraint preventing the minority from taking advantage of government policies and programs (see World

[2]Without estimating how language barriers affect ethnic inequality, Baulch et al. (2010) do suggest that not being able to speak Vietnamese substantially increases the minority's likelihood of being poor.

[3]For example, the work of Van de Walle and Gunewardena (2001) noted (but did not report) their attempt to use a language dummy in regressions for the minority group using the 1993 household data, and found no significant effects.

Bank, 2009a; Vasavakul, 2003; Tran, 2004; Vietnam Academy of Social Sciences, 2009, among others).[4] Our findings quantitatively corroborate this claim.

Second, our work explores the role of infrastructure in explaining the ethnic gap. There are two reasons for the need to consider infrastructure in quantifying the expenditure gap by ethnicity. The first is that the minority tends to live in more remote areas, characterised by difficult terrain, poor roads, no power and limited access to markets, making it difficult to isolate the effect of ethnicity itself on the expenditure gap. The second reason is that majority households living in the same remote and impoverished areas are doing increasingly well, to the point of being difficult to distinguish from their counterparts in the low-land communes (Swinkels and Turk, 2006). Recent literature indeed suggests that majority households benefit more from local investment and government poverty reduction programs than other groups (Pham et al., 2011; World Bank, 2009a, p.3). Our result challenges this view and generally supports the need for infrastructure improvement programs by the Vietnamese Government for the minority.

Despite a number of different economic studies on ethnic inequality in Vietnam, using various estimation techniques, our paper differs substantially in terms of estimation method. Typically, economic studies on ethnic inequality in Vietnam have tended to focus on differences in the household-specific characteristics, including differences in demographic structure, education and land, along with returns to those characteristics, to explain expenditure gaps among ethnic groups (see Van de Walle and Gunewardena, 2001; Baulch et al., 2007; Hoang et al., 2007; Baulch et al., 2012, 2010). This objective is complicated by two potential concerns. The first is the existence of common commune-specific unobserved characteristics, such as local customs, practices, and land and school quality, which are likely correlated with household-specific characteristics. Failure to control for this correlation, for example, in the use of an OLS estimator (e.g., Baulch et al., 2012, 2010), causes potential bias in estimating returns to those household-specific characteristics (Hsiao, 2003; Baltagi, 2005). This bias can be eliminated by using least-squares dummy-variable (LSDV) or fixed effects (FE) estimators, the commonly used approach in the earlier literature (e.g., Van de Walle and Gunewardena,

[4]For cross-country comparisons, Grafton et al. (2007) find that language barriers generate social barriers to communication and impede knowledge transfer and productivity. In various country studies, Patrinos et al. (1994) and Parker et al. (2005) report school inequality and language barriers for indigenous children, and Chiswick (1991), Chiswick and Miller (1995), Chiswick et al. (2000) and Dustmann and Fabbri (2003), among others, show evidence of the importance of language skills in labour market participation and the earnings of immigrants.

2001; Hoang et al., 2007). However, this way of eliminating the bias comes at the expense of the ability to estimate impacts of commune-specific observed attributes such as geographical characteristics and infrastructure.

The second concern is reverse causation. A good example of this occurs in cases where household expenditure patterns determine the provision of commune infrastructure, rather than the reverse. Ignoring this possibility could potentially lead to bias in estimating returns to infrastructure.

In this paper, we apply an instrumental variable approach to address both of these concerns. Our estimators of both household-specific and commune-specific observed variables are consistent and avoid potential bias while still capturing commune-observed effects.

The paper is organised as follows. Some background is provided in Section 2, including detail on ethnicity and the key government programs designed to assist the minority. Data and variables are described in Section 3. Section 4 provides the model specification and estimation method and section 5 presents results. Section 6 concludes, highlighting policy implications and scope for further research.

2.2 Ethnicity, Migration and Programs Affecting the Minority

According to the official classification of the Government of Vietnam, there are 54 ethnic groups living in Vietnam, including the Kinh and Chinese (Bui, 1999). The Kinh group accounts for about 86 per cent of the population. In spite of their diversity, ethnic minorities are usually grouped based on the place where they live. For instance, there is a tendency to lump together the ethnic groups in the Northern Mountains, in the Central Highlands, and in the lowlands. The largest proportion (about 12 million people) lives in the first two areas. The group in the lowlands comprises mainly the Chinese in Ho Chi Minh City, the Khmer in the Mekong Delta and the Cham in the Southern Coast. The biggest groups after the Kinh are the Tay (1.2 million people), the Thai (one million), and the Khmer (one million). The smallest groups, including the Si La, the Pu Pep, the Ro Man, the Brau and the O Du, include less than one thousand people each (Huynh et al., 2002).

Ethnic minorities have been affected by the consequences of Doi Moi, the process of economic reform and trade liberalisation initiated in 1986, much the same as the rest of the population. These economic reforms have resulted in an unambiguous improvement in living standards in Vietnam. A combination of better incentives, improved access to markets and government support was critical to this success. But ethnic minorities have also been affected, not always positively, by specific public policies and programs. Some of those policies and programs have had a dramatic impact on their livelihoods, both before and after Doi Moi.

Until the beginning of the 20th century, the Kinh people largely lived in deltas and lowland coastal areas while other non-Vietnamese speaking ethnic communities occupied the upland mountain areas (World Bank, 2009a). The establishment of new economic zones resulted in a massive migration of Kinh people to areas that had been traditionally inhabited by ethnic minorities. Migrations were largely driven by economic considerations, such as the desire to develop mountainous areas or spread population more evenly across the country (World Bank, 2009a). As a result, majority households received more support from the Vietnamese Government to migrate into minority areas, with far less support available to ethnic minorities. The World Bank (2009a, p.27), for example, notes that "some investment programs for the highlands, particularly in the Central Highlands, initially focused on bringing in Kinh migrants to set up services and work opportunities with the State, rather than hiring or promoting local ethnic minorities."

These migration movements can be divided in three periods. Between 1960 and 1975, economic zones in the mountainous areas took the form of state agricultural and forest enterprises, as well as new economic villages. In total, some 920,000 people from the Red River Delta were resettled in the Central Highlands and the Northern Mountains areas, and 80,000 people in the coastal areas. Subsequently, between 1976 and 1986, the government established a planned migration program to support the development of state forests and farms. As a result, an additional 710,000 people moved to the Central Highlands and some 200,000 to the Northern Mountains. In 1987, the planned migration program slowed down, due to a shortage of funding. But spontaneous migration soared, with as many as 2.3 million people moving during the 1980s, and around 300,000 more every year from then on (Huynh et al., 2002).

These massive population movements affected access to land by ethnic minorities and even the ecosystems on which their livelihoods depended. For example, in the

Central Highlands, from 1975 to 2000, total forest area falls (with clearing) from 4 million to about 2.9 ha, and of the remaining forest, state forestry enterprises occupy as much as 50 percent of the available land. The local indigenous people which accounted for the most population in the Central Highlands in 1975 represented only 26 percent of the population in early 2000s (Luu, 2010). As much as 60 percent local indigenous households were reported as having no production land as of 2002, while most of the fertile land was in hand of immigrants (Luu, 2010). As a result, local indigenous people were pushed further into the forest to create new farm and/or to work for immigrants on the land they used to own.

Another government initiative directly affecting ethnic minorities was the so-called "sedentarisation program" launched in 1968. This program aimed to reduce poverty and eliminate hunger in the mountainous regions by providing support for agricultural production and livelihoods. The program facilitated fixed settlement and cultivation, while also providing assistance for technical training, capacity building, technology transfer and raising market awareness. By 1998, when activities of the "sedentarisation program" were merged into Program 135, for the socioeconomic development of the most disadvantaged communes, about 3.8 million people had been resettled.

Program 135 was one of the first programs to concentrate not on population movements, but on direct support for minorities. It was established in 1998 to improve the living standards of ethnic minorities in so-called "extremely difficult" communes, and to narrow the development gap among ethnic groups and regions throughout the country. With a commune-targeting approach, the program has largely financed infrastructure development in these troublesome communes, with the number of defined "extremely difficult" communes increasing from 1,200 in 1999 to 2,410 communes in 2005 (Committee for Ethnic Minorities, 2005), accounting for about 25 percent of total communes in Vietnam.

Roughly 90 percent of the funding for Program 135 came from the central state budget. Other funding sources included local budgets and mobilised funds from various sources. During Phase I from 1999 to 2005, the total investment fund of Program 135 was 10,178 billion VND (equivalent to about 650 million USD[5]). Investments focused on transportation (40 percent of total investment), schools (23 percent), irrigation (17 percent), electricity (8 percent), water supply (6 percent), and clinics (2 percent) (Committee for Ethnic Minorities, 2005).

[5]We use the prevalent exchange rate of VND 15,700 = 1 USD in 2005 for conversion.

The Hunger Eradication and Poverty Reduction Program (HEPR) was launched in 1998 with an objective to eliminate chronic hunger and reduce the percentage of poor households in the country. HEPR, together with a wide range of health and education exemption policies, specifically targeted households classified as hungry or poor, many of whom are from the minority group. HEPR, in particular, focuses on providing access to credit, exemption from education fees, and support for health care, among other benefits, to entitled households (see Nguyen and Baulch, 2007). Funding for HEPR comes mainly from the central state budget (about 75 percent), with support from local budgets (about 25 percent). From 2001 to 2005, total funding for HEPR was roughly 6,240 billion VND (equivalent to about 400 million USD) (Ministry of Labour, Invalids & Social Affairs and UNDP, 2004).

2.3 Data and Variables

2.3.1 Data

Our estimates rely on the Vietnam Household Living Standard Survey in 2006 (VHLSS 2006). This is a multi-stage stratified household survey on expenditure, among other indicators, representative of the whole country, for both urban and rural areas. Three households were randomly selected from a single census enumeration area in each commune, making up 9,189 households, living in 3,063 communes.

Since the survey does not cover all the different ethnic groups in full detail, we are forced to simplify the ethnic divisions we consider. This lack of information is compounded by the fact that both the poor and ethnic minority households tend to live in rural areas, where we are interested in the determinants of poverty for ethnic minorities alone. To focus attention on the questions of interest, and avoid undue confounding of factors, we follow previous research in limiting our sample to rural households, which covers most ethnic minority groups, and then group households into a majority group, combining Kinh and Chinese households, and a minority group, which includes the remaining 52 officially recognised ethnicities.

One important characteristic of many communes is the presence of majority and minority households within the same commune. However, precise information on actual ethnic breakdown in Vietnam's rural communes is not directly available in the survey. As a result, we use information on households surveyed per commune, together with an ethnicity indicator, to construct a proxy of the overall ethnic composition of the commune.

Expenditure data from VHLSS 2006 also does not allow us to get a good proxy of ethnic composition, since there are only three households surveyed per commune. We overcome this by supplementing with income data, which includes more households per commune. Based on these constructions, we describe communes as either mixed in the sense they include both the majority and the minority, or non-mixed, when all households sampled in a commune belong to the same ethnic category.

In total, the rural sample used for this paper has 5,392 majority households and 1,168 minority households in 2,187 communes. Using information from income data, 1,611 communes have observations only from the majority group and 243 communes have observations only from the minority. On the other hand, 560 majority households and 439 minority households are found in 254 and 214 mixed communes, respectively.

2.3.2 Variables and Summary Statistics

The key dependent variable we model is real household expenditure per person. It is measured in thousand Vietnamese Dong, using January 2006 prices to correct for different inflation rates due to different survey times in different regions of the country. It might be argued that income per person, rather than expenditure, is a better indicator of economic welfare, and thus the welfare gap between ethnic groups. Nevertheless, income is difficult to measure in a developing country like Vietnam, particularly for the rural and minority groups we are studying. In rural areas, there are often no well-developed input markets to compute net income from farming and household activities, and no reliable measures of 'own-income' for household-managed and operated farms, making it difficult to distinguish between revenue and costs (Che et al., 2006), not to mention the potentially large amount of 'informal' or unreported income. The extent of own and informal income might differ systematically between social groups, therefore further distorting measured income as an indicator of group welfare. Another advantage of using expenditure as a welfare measure is that consumption tends to be smoothed in response to income fluctuations over relatively a long period of time (Deaton, 1997), as thus may be a better indicator of economic welfare than a snapshot of income which is possibly highly transient.

The explanatory variables we use are characteristics and endowments at both the household and commune levels. These include the demographic characteristics of the household and endowments of the household in human and physical capital. At the commune level, endowments as infrastructure are similarly distinguished from geographical characteristics.

Summary statistics for all variables are shown in Table 2.1. The characteristics of the majority and minority ethnic groups are often very different, and there clearly exist substantial gaps in endowments between the two groups. Of particular interest in this study, along with differences in language, infrastructure and years of schooling, real expenditures per person among the minority are more than a third lower than those for the majority group. The mean difference of the two groups is statistically different at the 1 percent level for all of the variables in Table 2.1, except for perennial land, other land, and existence of a primary school.

2.3.2.1 Household-level variables

Household characteristics of interest include household size, proportions of children in the 0-6 and 7-16 year brackets, proportions of male and female adults, household structures (describing whether a household consists of two generations with fewer than 3 children, two generations with three or more children, and three generations and other household structures), and some household-head specific variables such as age and gender. Table 2.1 suggests that the minority is more likely to have larger families, live in three-generation households and be headed by a man, while majority households tend to have a higher proportion of members over 16.

Household characteristics are expected to have an impact on household expenditure per capita. Economies of scale and complementarity in consumption and production are all factors evident in literature (Deaton and Muellbauer, 1980b; Lazear and Michael, 1988), suggesting the need to control for household size and its composition while other explanatory variables are being considered. A pattern of strong negative correlation between household size and consumption per person is found in many household surveys, especially in developing countries (Lipton and Ravallion, 1995). Increasing returns in household production due to specialisation

TABLE 2.1: Descriptive Statistics

Variables	Majority Mean	Minority Mean	Difference Mean
Real per capita expenditure in Jan. 06 prices (VND000s)	5151	2916	2235***
Household Characteristics			
Household size	4.10	5.14	-1.04***
Proportion of household member(s)			
Aged from 0 to 6	0.07	0.12	-0.05***
Aged from 7 to 16	0.19	0.24	-0.05***
Males, aged over 16	0.34	0.31	0.03***
Females, aged over 16	0.39	0.33	0.06***
Proportion of households that consist of			
Parent(s) only/Parent(s) with fewer than 3 children	0.54	0.36	0.18***
Parent(s) and three or more children	0.21	0.30	-0.09***
Three-generation households	0.13	0.17	-0.04***
Other household structures	0.12	0.17	-0.05***
Headed by female (yes = 1)	0.22	0.11	0.11***
Age of household head	49.93	43.95	5.98***
Household Human Capital:			
Require interpretation	0.00	0.28	-0.28***
Years of schooling of the most educated member	9.58	8.01	1.57***
Household Physical Capital: Land area (1000m2)			
Irrigated annual land	2.74	3.24	-0.50**
Non-irrigated annual land	0.48	4.93	-4.45***
Perennial land	1.42	1.48	-0.06
Forestry land	0.61	5.54	-4.93***
Water-surface land	0.39	0.11	0.28***
Other land	0.03	0.07	-0.04
Commune Characteristics:			
Proportion of households living in communes located in			
Rural coastal and delta land	0.72	0.10	0.62***
Rural hilly land	0.08	0.01	0.07***
Rural low/high mountains	0.20	0.89	-0.69***
Distance to city (1000 km)	0.14	0.25	-0.11***
Commune Infrastructure:			
Proportion of households living in communes that have			
Power	0.99	0.95	0.04***
Clean water	0.52	0.13	0.39***
Hard-surfaced car road	0.47	0.20	0.27***
Primary school	0.50	0.52	-0.02*
Village daily market	0.36	0.11	0.25***
State-owned enterprise within 10km	0.21	0.13	0.08***
Foreign-shared enterprise within 10km	0.13	0.02	0.11***
Local enterprise	0.67	0.40	0.27***
Number of households	5392	1168	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

and complementarity of skills could also be expected in agricultural and household businesses, while differences in the age and gender structure could influence consumption patterns across households. Furthermore, as child labour is not unusual in Vietnam (Edmonds and Pavcnik, 2005), having an additional household member beyond 7 years old in the household may reduce the total 'time cost' per person for cooking and cleaning, and hence free time for other members to work.

Household human capital in our data consists of two variables. The first variable is the years of schooling of the most educated member. Schooling has been widely shown to have strongly positive impacts on income, and hence on expenditures. The previous literature on ethnic inequality in Vietnam measured schooling as the highest level of attainment, and used indicator dummy variables to show the impact of different education levels on household living standards. We believe that schooling years as a (nearly) continuous variable will allow us to measure incremental returns to education in a better way. In our data, the majority group dominates with an average of 9.6 years of schooling of the most educated member, significantly more than the minority group with about eight years.

The second variable on household human capital of interest is fluency in Vietnamese. In the survey, this is simply indicated by whether an interpreter was required to complete the survey. Language barriers are expected to influence the household living standard of the minority. For example, the minority often reports on their lack of knowledge of government policies and programs due to their inability to read or hear about these measures, let alone to request government services they are entitled to (World Bank, 2009a). Language barriers prevent minority women, in particular, from using government health services, even though they possess health care coverage cards (Tran, 2004). This barrier also results in a higher likelihood of minority children beginning school late, repeating school, or dropping out, compared with their majority counterparts (World Bank, 2009a). Lack of language skills, which leads to lack of confidence and limits social networking, also hinders the minority from sharing information, and accessing off-farm employment outside their specific group (World Bank, 2009a; Vietnam Academy of Social Sciences, 2009). In our sample, the minority group has a significant language barrier, as shown in Table 2.1, with 28 percent of such households requiring the needed of language interpretation for the survey interview.

The primary measure of household physical capital in the survey is land. We include different types of land: irrigated annual, non-irrigated annual, perennial, forestry, water-surfaced land and other land. Land disaggregation by type is needed to control for heterogeneity in land productivity of various types of land, crops that they can produce, and land tenures, as well as their usefulness as collateral for loans (Kompas et al., 2012). In general, the quantity of land area is markedly in favour of minority households, except for the water-surface land where holdings by majority households are nearly four times as great as minority households. The distinction is important because of great variability in land quality. It is generally the case that the more fertile and higher market-valued land is in the deltas and coastal areas.

2.3.2.2 Commune-level variables

In our model, the commune-level variables include commune characteristics and infrastructure. Commune characteristics capture inherent location effects. There are three types of communes in the data set: coastal and delta land, hilly land, and low and high mountain land. Distance from the commune to the nearest city is also a key identifier. Geography and distance to a city has been widely shown to be important in development (e.g., Audretsch and Feldman, 1996). There is sharp contrast between the proportions of households living in various types of communes: 72 percent of majority households are concentrated in rural coastal and delta land areas, while as much as 89 percent of minority households reside in rural low or high mountain areas. Likewise, minority households live almost twice as far from the city compared to majority households. All of this suggests the need to control for these commune characteristics in model specification.

Commune infrastructure reflects the existence of basic infrastructure, trade facilities and off-farm employment opportunities. We include four measures of basic infrastructure which cover access to power, clean water,[6] hard-surfaced roads, and a primary school.[7] Each of these variables has a plausible causal relationship with household income and expenditure. For example, one might expect having access to power would enhance household production and consumption and thus positively influence household expenditure. Similarly, clean water promotes

[6]Clean water includes water from the following sources: private tap water inside and/or outside the house, public tap water, water pumped from a deep drilled well, and bought water in bottles or tanks.

[7]We could not include the presence of health clinics in our research, despite its presence in Government Program 135, since as many as 98 percent of both groups have access to a health clinic, leading to little contrast in the data.

good health, which enhances economic productivity, and easy water access reduces labor-time and costs in transporting water. The availability of a hard-surfaced road greatly assists the transportation of goods and services, and reduces travel costs of dwellers both in time and money. Finally, the existence of a primary school would likely help increase community-wide education and enhance living standards.

We measure 'trade facilities' of households in the commune by the existence of a daily village market. Clearly, a daily village market is likely to help facilitate trade and the exchange of products, which helps promote household production and consumption.

Off-farm employment is often seen as one important channel out of poverty in rural areas. Its importance is quantified by three indicators in the survey in terms of the existence of different types of enterprises: state-owned enterprises within 10 km of the commune, foreign-shared enterprises within 10 km of the commune, and other enterprises located in the commune. Generally speaking, state-owned enterprises are often larger and reportedly less efficient, foreign-shared enterprises vary considerably in size and employment opportunities, and local enterprises, although often small in size, usually indicate the presence of a highly-developed local market economy.

For all items in commune infrastructure, minority households are often far less well off than their majority counterparts. An exception is access to a primary school where the difference by ethnicity is in favour of the minority, though marginal. Furthermore, many more majority households live in communes with alternative enterprises, of varying ownership types. One limitation of our data is that variables on commune endowments do not measure quality, or differences in quality, which are likely to vary across regions. If anything, we suspect the data we do have on infrastructure understates the differences between the two groups.

2.4 Model Specification and Estimation Method

Household expenditure per capita is assumed to be a function of household-specific and commune-specific characteristics and endowments. The dependent variable appears in the equation as its logarithm, representing constant proportional effects of explanatory variables on per capita household expenditure. Household size as an explanatory variable also appears in log form. We are interested in exploring how the two ethnic groups differ in their responses to their household-specific characteristics and endowments, which also vary on average across communes, as well as to commune-specific characteristics and infrastructure, which are variable only between communes. Our basic specification for each group is:

$$\ln E_{ji} = X'_{ji}\beta + Z'_i\gamma + [\alpha_i + \epsilon_{ji}], \text{for } j = 1, \ldots, M_i; i = 1, \ldots, N \quad (2.1)$$

where E_{ji} is the per capita household expenditure for household j living in commune i, X_{ji} is a $K \times 1$ vector of household-specific explanatory variables, which include household characteristics and household human and physical capital, while Z_i is a $G \times 1$ vector of commune-specific explanatory variables, which include commune characteristics and commune infrastructure. These variables are described in detail in the previous section.

We note explicitly that X_{ji} varies over both household j and commune i, while Z_i varies over only commune i. The dimensioned parameter vectors β and γ, the commune-specific intercept α_i and the error term ϵ_{ji} are all unobserved. There are M_i households in each of the $i = 1, \ldots, N$ communes, making NM observations in total, where $M = \sum_{i=1}^{N} M_i$.

The composite error term in equation 2.1, formed by the commune-specific intercept α_i and the error term ϵ_{ji} in the square brackets, captures both the variation in coefficients across communes and the usual idiosyncratic error term. If this composite error term was uncorrelated with each of the explanatory variables, OLS would provide a consistent estimator. Furthermore, with the assumptions of homoskedasticity and independence of effects, a random-effects (RE) estimator would be consistent and efficient. However, such an absence of correlation is unlikely.

The principal concern is correlation of the commune-specific unobserved effect α_i with commune averages of the household-varying explanatory variable X_{ji}. In our application, when households are sampled by communes, they share common commune-level effects, thereby forming a panel-like data structure. These commune-level effects partially reflect local practices and customs as well as the unobserved qualities of institutions and endowments such as schools and land. It is likely that these terms are correlated with the commune averages of household characteristics and endowments, such as family size, educational attainment and land ownership.

When the group effects are correlated with explanatory variables, OLS (and RE) produces inconsistent estimates for the coefficients β and γ. The standard remedy is to use a FE estimator, which is equivalent to including a dummy variable for each commune. Unfortunately, under the fixed-effects approach, coefficients γ on the commune-invariant variables Z are not identified.

In a model with a random intercept, Hausman and Taylor (1981) observe that the fixed-effects estimator can be viewed as an instrumental variables (IV) procedure, where the instruments are deviations in the X explanatory variables around their group means (the "within-group" variation). This interpretation suggests that the coefficients γ of the group-invariant variables Z can then be estimated, provided sufficient additional instruments to separately identify γ are available. Some commune characteristics, such as geography and distance to the city, denoted as Z_1, can clearly be seen as exogenous. In such cases, these variables can be instrumented by themselves (Breusch et al., 2011).

By contrast, commune infrastructure, denoted as Z_2, may not be exogenous. The concern is that the availability of infrastructure is jointly determined with the dependent variable, per capita household expenditure. As described in section 2, provision of basic infrastructure including roads, power, water and primary schools is the backbone of Program 135, which supports poor and ethnic minority communes. For this reason, provision of this basic infrastructure would be influenced in part by commune (average) living standards and the ratio of minority households in the communes, which is, itself, also often correlated with the commune average living standard. Also, the provision of basic infrastructure is often based on economies of scale: how easy it would be to provide infrastructure, how many households can benefit from it, and to what extent it would be used. To this end, highly populated communes with easy access, or those expected to have higher living standards on average, would more likely attract public investment. Finally, the establishment of enterprises and daily markets would depend on basic commune infrastructure, as well as average labour skills and production capacity.

We address these concerns by constructing instrumental variables that remove possible causal links from household expenditures to commune infrastructure Z_2. To do so, following Hausman and Taylor (1983), we use as instruments for commune endowments Z_2 the residuals from equations predicting each of the Z_2 variables. The regressors in these prediction equations (below called H) include an indicator of whether the commune is under Program 135, as well as the ratio of minority households, commune population size and commune averages of all household human and physical capital. These instruments are particularly attractive because they pick out the variation in household expenditure caused by commune infrastructure Z_2, rather than a causal link from the household expenditure to commune infrastructure Z_2.

To describe our IV estimator, some notation is helpful. We can stack the data with all households in all communes, so the model for NM observations can be written compactly as:

$$Y = X\beta + (Z_1)\gamma_1 + (Z_2)\gamma_2 + \epsilon \tag{2.2}$$

It is useful to define the $NM \times NM$ "within" operator Q_C as the matrix that converts any variable of data, such as $Y = \ln E$ and the columns of X, into the deviations from its commune-level means. In this notation, the instruments for the household characteristics X are $Q_C X$, their within-commune deviations. If these were the only available instruments, the estimator would be a simple FE estimator and the coefficients γ would remain unidentified.

Given the regressors H in the prediction equations that are used to form the additional instruments, a projection matrix P_H can be defined such that $(I-P_H)Z_2$ is the matrix of residuals that provides the instruments. The IV estimator has the following moment conditions:

$$[Q_C X, Z_1, (I-P_H)Z_2]'[Y - X\beta - Z_1\gamma_1 - Z_2\gamma_2] = 0, \tag{2.3}$$

which decompose as:

$$\begin{aligned} &[Q_C X]'(Y - X\beta) = 0\ , \\ &[Z_1, (I-P_H)Z_2]'[Y - X\beta - Z_1\gamma_1 - Z_2\gamma_2] = 0 \end{aligned} \tag{2.4}$$

The first of these conditions in 2.4 shows that the estimator of β is the same as fixed-effects. In particular, it will be consistent in situations when there exists correlation between the group-specific random intercept α_i and a group-varying explanatory variable X_{ji}. The second condition in 2.4 assumes that Z_1 is exogenous and that the residual from equations predicting each of Z_2 on H is uncorrelated with the error term (hence it is a valid instrument). When these conditions are satisfied, the estimators of γ_1 and γ_2 will be consistent.

One of the key concerns in our research is the difference in returns to characteristics and endowments between groups. We could estimate separate regressions specified in equation 2.1 for majority and minority sub-samples to identify the average coefficients for each group. Instead and equivalently, we add interactions between the covariates and group dummy-variables to the same regression applied to the whole sample, which has the advantage of providing standard errors of the difference in coefficients between groups.

2.5 Estimation Results

We first check if our estimation method is appropriate, and then examine the effect of language, mixed communes and infrastructure, with a focus on differences in returns to those attributes between the two ethnic groups. Finally, after various robustness checks, we decompose the ethnic gap.

2.5.1 Checking the Estimation Method

We begin by testing the need for using the first set of instruments, the within-commune deviations $Q_C X$ for household-specific characteristics X. The Hausman test, which compares coefficients estimated using RE and FE estimators, rejects the null hypothesis that a RE estimator is consistent ($\chi^2_{25} = 140.34$), suggesting the need for a FE estimator. This test result confirms our concern about the correlation between commune-level unobserved effects and (commune-level averages of) household-level characteristics and endowments, thereby lending credence to our use of within-commune deviations $Q_C X$ as instruments.

We next consider if commune-level infrastructure Z_2 is possibly endogenous. The concern is that infrastructure is jointly determined with (commune-level averages of) household expenditure, and so its use as a regressor will fail to identify the extent that households benefit from having access to infrastructure. In our data, there are relatively strong correlations between household expenditure and the existence of Program 135, and ratios of minority households in the commune and commune population sizes (-0.27, -0.40 and 0.27, respectively). Furthermore, as shown in Tables 2.2 and 2.3, regressions explaining the availability of each Z_2 by the variables in H are all (and overall) significant. Indeed, Program 135, commune population size, and the ratio of minority households are highly significant in most regressions, suggesting their strong influence on the availability of commune infrastructure. Although such regressions do not prove that these infrastructure variables Z_2 are endogenous, the regression results do support our concerns about possible endogeneity and endorse our strategy for constructing suitable instrumental variables.

Unlike simple FE estimation, where the nature of the instruments obviates the need to adjust for the error correlation that is due to the panel structure of the data, in this broader IV estimation the error correlation should be taken into account. We conduct this IV estimation in Stata, version 11, using the command *ivreg2* with option *vce(cluster commune)* to correct for cluster correlation at the commune level and heteroskedasticity. The F-statistic for the first stage regression for the whole rural sample is 34.64. It is well above the critical value of 21.42, identified by Stock and Yogo (2005) for a 5 percent maximal IV relative bias. It thus appears to exclude a possible problem with weak instruments.

Finally, we perform a Chow test for whether the coefficients estimated over the majority are equal to the coefficients estimated over the minority group. The tests soundly reject the null hypothesis that they are equal between groups (χ^2_{28} = 192.51).

2.5.2 The Effect of Language

Table 2.4 reports results of estimating equation 2.2 by an IV estimator with the instruments described in equation 2.3. We find a negative and significant effect of the presence of a language barrier on household expenditure. Being language 'incompetent' – requiring interpretation during the survey interview – results in an approximately 14 percent fall in household expenditure in the pooled model. This negative result is not sensitive to the choice of instruments for language ability, which we will discuss in detail in our robustness checks. Furthermore, this impact remains almost the same in the regression for minority households only, suggesting that even among minority households, where people can communicate in their own minority language, not being fluent in Vietnamese imposes a substantial disadvantage.

TABLE 2.2: Regressions Predicting Z_2 by H for Minority Group

	(1) Power	(2) Clean water	(4) Hard surfaced car road	(5) Primary school	(6) Daily market	(7) Local enterprise	(8) Foreign shared enterprises	(9) State-owned enterprise
Program 135	-0.049***	0.095***	-0.027	0.053*	-0.011	-0.074**	-0.033***	-0.065***
(yes=1)	(-3.57)	(5.72)	(-1.07)	(1.68)	(-0.60)	(-2.51)	(-3.33)	(-3.09)
Commune population size	0.022***	0.234***	0.081***	-0.033*	0.114***	0.142***	0.006	0.069***
(1000 hholds)	(2.76)	(24.79)	(5.68)	(-1.85)	(10.55)	(8.47)	(1.13)	(5.69)
Ratio of minority households	0.020	-0.104**	-0.023	-0.095	-0.156***	-0.186**	-0.036	-0.143**
sampled in the commune	(0.55)	(-2.33)	(-0.34)	(-1.11)	(-3.06)	(-2.35)	(-1.37)	(-2.52)
Years of schooling	0.007***	-0.002	0.001	-0.023***	0.005	0.018***	-0.000	0.014***
	(2.78)	(-0.62)	(0.18)	(-3.86)	(1.45)	(3.35)	(-0.07)	(3.63)
Irrigated crop land	-0.002**	0.001	0.003	0.003	0.001	-0.003	0.000	0.001
(1000 m2)	(-2.30)	(0.79)	(1.56)	(1.46)	(0.98)	(-1.28)	(0.03)	(0.81)
Non-irrigated crop land	-0.006***	-0.001	0.002	0.001	-0.002	-0.006***	-0.000	0.001
(1000 m2)	(-7.61)	(-1.01)	(1.61)	(0.41)	(-1.49)	(-3.70)	(-0.56)	(0.96)
Perennial land	0.002	-0.005***	-0.001	0.003	-0.002	-0.002	0.001	0.003
(1000 m2)	(1.33)	(-3.75)	(-0.29)	(1.15)	(-1.01)	(-0.94)	(0.93)	(1.57)
Forestry land	0.000*	-0.000	0.000	-0.001	-0.001*	-0.000	-0.000	-0.000
(1000 m2)	(1.65)	(-1.29)	(0.78)	(-1.06)	(-1.91)	(-0.36)	(-0.97)	(-0.91)
Water-surface land	-0.000	0.017***	0.005	-0.020*	0.005	-0.012	-0.001	-0.004
(1000 m2)	(-0.03)	(2.94)	(0.63)	(-1.79)	(0.83)	(-1.19)	(-0.41)	(-0.54)
Other land	0.004	-0.008	-0.011	0.008	-0.007	0.003	-0.001	-0.006
(1000 m2)	(0.53)	(-0.91)	(-0.86)	(0.48)	(-0.77)	(0.18)	(-0.26)	(-0.54)
Constant	0.904***	-0.083	0.110	0.789***	0.099	0.342***	0.072**	0.100
	(19.02)	(-1.45)	(1.27)	(7.18)	(1.51)	(3.37)	(2.13)	(1.37)
Number of households	1168	1168	1168	1168	1168	1168	1168	1168
F	14.867	81.615	4.677	3.847	19.769	18.493	2.590	9.885
r2	0.114	0.414	0.039	0.032	0.146	0.138	0.022	0.079

Except for Program 135, commune population size, and ratio of minority households sampled in the commune, all variables are commune averages. t ratio in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 2.3: Regressions Predicting Z_2 by H for Majority Group

	(1) Power	(2) Clean water	(4) Hard surfaced car road	(5) Primary school	(6) Daily market	(7) Local enterprise	(8) Foreign shared enterprises	(9) State-owned enterprise
Program 135	0.001	0.043*	-0.108***	0.095***	0.028	-0.173***	-0.052***	-0.072***
(yes=1)	(0.63)	(1.85)	(-4.59)	(4.04)	(1.24)	(-7.88)	(-3.25)	(-3.71)
Commune population size	0.001*	0.087***	0.035***	0.064***	0.069***	0.098***	0.015***	0.022***
(1000 hholds)	(1.92)	(12.27)	(4.85)	(8.84)	(10.01)	(14.54)	(3.09)	(3.65)
Ratio of minority households	-0.029***	-0.406***	-0.330***	0.011	-0.173**	0.074	-0.101*	0.185***
sampled in the commune	(-7.88)	(-5.38)	(-4.32)	(0.14)	(-2.35)	(1.05)	(-1.94)	(2.95)
Years of schooling	0.000	-0.004	0.010***	-0.016***	0.003	0.013***	0.010***	0.016***
	(0.62)	(-1.17)	(3.27)	(-5.18)	(0.90)	(4.52)	(4.80)	(6.33)
Irrigated crop land	0.000	0.002*	-0.005***	0.003***	-0.004***	-0.003**	-0.003***	-0.003***
(1000 m2)	(0.49)	(1.86)	(-4.51)	(2.81)	(-3.62)	(-2.42)	(-3.99)	(-3.59)
Non-irrigated crop land	-0.000***	-0.010***	-0.009***	0.000	0.005**	0.000	-0.002	-0.001
(1000 m2)	(-4.04)	(-4.26)	(-3.85)	(0.16)	(2.21)	(0.06)	(-1.10)	(-0.42)
Perennial land	-0.000	-0.006***	-0.003**	0.001	-0.000	-0.000	-0.001	-0.001
(1000 m2)	(-0.65)	(-5.79)	(-2.39)	(1.06)	(-0.32)	(-0.01)	(-1.01)	(-1.45)
Forestry land	0.000	-0.002***	-0.001	-0.001**	0.000	-0.000	-0.001	0.001*
(1000 m2)	(0.46)	(-3.07)	(-1.56)	(-2.10)	(0.19)	(-0.61)	(-1.12)	(1.94)
Water-surface land	0.000	0.009***	0.001	0.003*	-0.002	-0.005***	-0.002	-0.002
(1000 m2)	(0.02)	(4.67)	(0.48)	(1.80)	(-1.13)	(-2.86)	(-1.63)	(-1.24)
Other land	0.000	0.010	0.011	0.007	-0.005	0.006	-0.000	0.015**
(1000 m2)	(0.24)	(1.19)	(1.27)	(0.81)	(-0.54)	(0.74)	(-0.06)	(2.00)
Constant	0.998***	0.375***	0.341	0.495***	0.194***	0.356***	0.020	0.025
	(582.21)	(10.54)	(9.45)	(13.70)	(5.59)	(10.61)	(0.82)	(0.85)
Number of households	5392	5392	5392	5392	5392	5392	5392	5392
F	9.200	30.979	14.888	16.056	12.311	32.360	8.079	9.835
r2	0.017	0.054	0.027	0.029	0.022	0.057	0.015	0.018

Except for Program 135, commune population size, and ratio of minority households sampled in the commune, all variables are commune averages. t ratio in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 2.4, column (4), highlighting differences between groups, and where language differences are excluded, we see little evidence of significant difference in returns to characteristics and endowments. An exception is in years of schooling and to a lesser extent, in water-surface land. This is in spite of substantial difference in intercepts, as shown in column (1), where being a minority household is associated with a roughly 24 percent reduction in expenditure.

To further investigate the effect of language in household returns to attributes, we compare returns between households with and without language ability (i.e., those that do not or do require interpretation). In Table 2.5, for brevity, we present only variables with statistically significant difference in returns. A key finding is that the difference in returns to education between these two language groups is highly significant. This difference is much larger than that between the two ethnic groups: 0.041 per year of schooling as compared with 0.024. Language barriers thus clearly contribute to a widening of the ethnic gap, at least through the channel of differences in household gains from education.[8]

Does the ethnic gap dissipate among language competent households? In the regression for language competent households, we measure this gap using an indicator of 'belonging to the minority group'. By doing so, we assume that the returns to attributes are the same among language competent households. Results are shown in Table 2.5, where we find the ethnic gap remains almost the same as in the pooled regression. This implies that being competent in Vietnamese is not enough to eliminate the ethnic gap in expenditures.

[8]World Bank (2004) finds that students who always spoke Vietnamese outside school or belonged to the ethnic majority Kinh group were likely to have higher test scores than students who never speak Vietnamese outside school, or who belonged to the ethnic minority group.

TABLE 2.4: Difference in Determinants of Household Expenditure by Ethnicity

	(1) Pool	(2) Majority	(3) Minority	(4) Difference
Household Characteristics				
Household size (log)	-0.333***	-0.327***	-0.392***	0.065
	(-12.99)	(-11.61)	(-6.33)	(0.96)
Aged from 7 to 16	0.158***	0.151***	0.298***	-0.147
	(3.21)	(2.66)	(2.82)	(-1.22)
Males, aged over 16	0.482***	0.476***	0.571***	-0.095
	(8.31)	(7.26)	(4.30)	(-0.65)
Females, aged over 16	0.424***	0.425***	0.507***	-0.082
	(6.69)	(6.05)	(3.11)	(-0.47)
Parent(s) and three/more children	-0.060***	-0.068***	-0.028	-0.040
	(-3.54)	(-3.52)	(-0.75)	(-0.98)
Three-generation households	-0.056***	-0.052**	-0.058	0.006
	(-2.68)	(-2.22)	(-1.20)	(0.11)
Other household structures	0.000	0.013	-0.024	0.037
	(0.02)	(0.49)	(-0.45)	(0.62)
Headed by female	-0.071***	-0.069***	-0.086*	0.017
(yes = 1)	(-4.11)	(-3.69)	(-1.93)	(0.35)
Age of household head	0.011***	0.014***	-0.002	0.016**
	(3.26)	(3.65)	(-0.31)	(1.97)
Household Human Capital:				
Require interpretation	-0.141**	.	-0.120**	.
	(-2.57)	.	(-2.23)	.
Years of schooling	0.054***	0.058***	0.034***	0.024***
	(24.06)	(23.49)	(5.79)	(3.60)
Household Physical Capital:				
Irrigated annual land	0.009***	0.010***	0.007**	0.003
(1000 m2)	(5.73)	(6.91)	(2.37)	(0.96)
Non-irrigated annual land	0.006***	0.004	0.009***	-0.005
(1000 m2)	(3.28)	(1.47)	(4.38)	(-1.21)
Perennial land	0.009***	0.008***	0.013***	-0.005*
(1000 m2)	(7.35)	(6.59)	(5.34)	(-1.75)
Forestry land	0.001***	0.001**	0.001**	0.000
(1000 m2)	(2.84)	(2.38)	(2.20)	(0.12)
Water-surface land	0.012***	0.011***	0.021***	-0.010**
(1000 m2)	(4.20)	(3.97)	(5.67)	(-2.17)

	(1) Pool	(2) Majority	(3) Minority	(4) Difference
Other land	0.006*	0.007*	-0.003	0.010
(1000 m2)	(1.81)	(1.88)	(-0.14)	(0.48)
Commune Characteristics:				
Rural low or high mountains	-0.039**	-0.014	-0.092	0.078
	(-1.98)	(-0.66)	(-1.35)	(1.10)
Rural hilly land	0.015	0.015	0.060	-0.045
	(0.51)	(0.51)	(0.54)	(-0.39)
Distance to city	-0.287***	-0.335***	-0.224*	-0.111
(1000 km)	(-5.04)	(-5.30)	(-1.81)	(-0.81)
Commune Infrastructure:				
Power	0.170***	0.142	0.156**	-0.014
	(2.85)	(1.01)	(2.40)	(-0.16)
Clean water	0.031**	0.027*	0.123*	-0.096
	(2.06)	(1.76)	(1.94)	(-1.49)
Hard-surfaced car road	0.004	-0.006	0.072**	-0.078**
	(0.28)	(-0.37)	(1.99)	(-1.99)
Primary school	0.028**	0.036**	-0.005	0.041
	(2.02)	(2.35)	(-0.17)	(1.24)
Village daily market	0.048***	0.047***	0.072	-0.025
	(3.23)	(3.01)	(1.50)	(-0.50)
State-owned enterprise	0.021	0.018	0.051	-0.033
(within 10km)	(1.28)	(1.03)	(1.10)	(-0.66)
Foreign-shared enterprise	0.091***	0.092***	0.003	0.089
(within 10km)	(3.95)	(3.88)	(0.03)	(1.13)
Local enterprise	0.061***	0.058***	0.072**	-0.014
	(4.23)	(3.68)	(2.16)	(-0.37)
Minority Group	-0.242***	.	.	.
	(-9.00)	.	.	.
Constant	7.281***	7.435***	7.752***	-0.317
	(71.81)	(44.33)	(40.51)	(-1.17)
Number of households	6560	5392	1168	6560
Number of communes	2187	1865	457	2187

Log of real per capita expenditure as dependent variable. t statistics are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression omits the proportion of members aged 06; one- or two-generation households that consist of no more than two children; commune geographical types as coastal/delta; other dummy variables having yes=1 and no=0.

TABLE 2.5: Difference in Determinants of Household Expenditure by Language Ability

	(1) Pool	(2) Not Require Interpretation	(3) Require Interpretation	(4) Difference
Years of schooling	0.054***	0.056***	0.015	0.041***
	(24.06)	(24.60)	(1.12)	(2.92)
Foreign-shared enterprise	0.091***	0.091***	-0.368**	0.459***
(within 10km)	(3.95)	(3.96)	(-2.55)	(3.14)
Minority Group	-0.242***	-0.240***	.	.
(yes=1)	(9.00)	(10.57)	.	.
Require interpretation	-0.141**	.	.	.
(yes=1)	(-2.57)	.	.	.
Number of households	6560	6228	332	6560
Number of communes	2187	2099	136	2187

Pool regression and notes are the same as those in Table 2.4.

To further disentangle the ethnic gap among language 'competent' households, we narrow our sample to households that speak the majority language, and we relax our assumption that returns to attributes by ethnicity are the same. In Table 2.6, for brevity, we again only present variables with statistically significant differences between the two ethnic groups. In terms of returns per year of schooling, the results still favour the majority, albeit at now much smaller values. Specifically, the difference between groups falls to 0.014 per year of schooling, and is significant only at 10 percent level. This difference is less than two thirds of the difference between the two ethnic groups in the whole rural sample. On the other hand, the ethnic difference in returns to perennial land still favours minority households and becomes almost twice as much as that in the whole rural sample. Working on perennial land is a comparative advantage of minority households, given their indigenous knowledge. Having language skills seems to further enhance this advantage. Results in Table 2.6, thus suggest, once again, that the ethnic gap could be narrowed by removing the language barrier through enhanced returns to education and (as well) perennial land of the minority.

2.5.3 The Effect of Mixed Communes

Table 2.7 extends the results in Table 2.6 by separating language 'competent' households by both ethnicity and whether they live in a mixed or non-mixed commune. The idea is to see if living in the same commune helps further reduce the ethnic gap after controlling for language. In this case, we find that the difference

TABLE 2.6: Difference in Determinants of Household Expenditure by Ethnicity among Households not Requiring Interpretation

	(1) Not Require Interpretation	(2) Majority	(3) Minority	(4) Difference
Years of schooling	0.056***	0.058***	0.044***	0.014*
	(24.60)	(23.49)	(6.49)	(1.92)
Perennial land	0.009***	0.008***	0.018***	-0.009***
(1000 m2)	(7.50)	(6.59)	(6.03)	(-2.89)
Minority Group	-0.240***			
(yes=1)	(-10.57)	.	.	.
Number of households	6228	5392	837	6228
Number of communes	2099	1865	356	2099

Regression for households not requiring interpretation is the same as that in Table 2.5. Notes are the same as those in Table 2.4.

in returns to education between majority and minority households in non-mixed communes is very small, or 0.009 per year of schooling and insignificant. On the contrary, the difference is significant and much larger, at 0.023 per year of schooling, between majority and minority households in mixed communes.

TABLE 2.7: Difference in Determinants of Household Expenditure by Ethnicity among Households not Requiring Interpretation in Mixed and Non-mixed Communes

	Non-mixed Communes		Mixed Communes	
	Majority	Minority	Majority	Minority
Years of schooling	0.059***	0.050***	0.061***	0.038***
	(22.73)	(6.72)	(7.67)	(3.67)
Perennial land	0.008***	0.012**	0.009***	0.018***
(1000 m2)	(6.06)	(2.13)	(3.01)	(4.31)
Number of households	4832	450	560	387
Number of communes	1611	163	254	193

Information based on ethnic mix is from Income data, thereby causing difference in the number of mixed communes for the majority and the minority. Difference in returns to years of schooling between majority and minority in mixed communes is significant (p=0.079); between majority and minority in non-mixed communes is insignificant (p=0.301); Difference in returns to perennial land between majority and minority in mixed communes is significant (p=0.073); between majority and minority in non-mixed communes is insignificant (p=0.476). Other notes are the same as for Table 2.4.

This larger difference in returns to education among language 'competent' households in mixed communes is initially surprising. In the absence of language barriers, one would expect to see that living in the same commune with the majority group would generate comparable returns to education for the minority. In mixed communes, one might also expect little difference in education quality — minority and majority students are in the same classrooms — and that the availability of employment in the commune to both groups should be comparable.

This wider disparity in returns to education suggests at least two possibilities. First, education quality may not be uniform to majority and minority groups in the same classroom. This will often be the case where special needs of minority children are not addressed in mixed classrooms, the way they might be in those with minority children only. In mixed classrooms, in other words, minorities may be 'left behind' compared to their majority counterparts. This will be especially the case if minority children enter the classroom with differential skills and prior differences in home-based learning.[9]

Second, the disparity in returns to education may suggest unequal treatment in labour markets within mixed communes. This is likely rooted in the government policy which has largely favoured Kinh movement into minority areas, rather than the reverse (World Bank, 2004; Huynh et al., 2002). As indicated in section 2, the labour market in mixed communes was distorted from migration patterns (World Bank, 2009a), which generally favoured the majority at the expense of minority groups in terms of concessions for migration and the establishment of preferential employment by local governments for the majority. If the majority is favoured in hiring in mixed communes — and differences between individuals by ethnicity are commonly known or identified at the point of the job application — this may explain the difference in returns and expenditure patterns. Indeed, the results show that that not only are returns to education of minority households in mixed communes much lower than those of the majority households, but that they are also much lower than those of their minority counterparts living in non-mixed communes.

2.5.4 The Effect of Infrastructure

In Table 2.4, we find no statistically significant difference in returns to commune infrastructure between the two ethnic groups. An exception is in returns to hard-surfaced road, which favours the minority. As such, this result offers us no evidence for the hypothesis, as recently argued in Vietnam, that majority households benefit

[9]There might be a concern that household heterogeneity may affect educational attainment and returns to education. Our use of within-commune deviations as an instrument for education does not correct for this possibility. However, if this bias is systematically similar for both the majority and minority groups, noting that we are interested only in the ethnic differences in educational attainment and returns to education, its effect would be small.

more than minority households from local investment (i.e., programs supporting poor and minority communes, such as Program 135). In fact, if anything, local investment seems to benefit minorities more.

It is also worth noting that even with no statistically significant difference, minority and majority groups exhibit relatively divergent patterns in their returns to infrastructure. This divergence is in contrast to the similarity in trend in returns to household physical and human capital. Understanding this divergence sheds light on what mechanisms outside the household could give rise to the living standard of the minority. There are a number of results to highlight in this regard.

First, our results show that the minority generally benefits more than the majority from basic infrastructure. Indeed, having access to power, clean water, and hard-surfaced roads increases household expenditure of the minority by 16 percent, 12 percent and 7 percent, respectively. On the contrary, the majority gains little, if anything, from these infrastructure facilities. Higher returns in poorer communes in rural settings have often been noted (e.g., Mu and Van De Walle, 2007). It is possible that poor communes, where minority households are often found to live, are generally underdeveloped, thereby offering greater scope for returns from infrastructure.

Second, it is clear from the evidence that primary schools are the only item among the four basic infrastructure categories (i.e. power, water, road and primary schools), largely funded by the Government, which benefits the majority more than the minority. Primary schools have no impact on the minority, but increase majority household expenditures by 3 percent. This may be an artefact of Vietnam's inability to provide a suitable language-based education for minority children. Generally speaking, Vietnam has no systematic teacher training for bilingual education (Vasavakul, 2003; World Bank, 2003b), and minority languages are not used as the main medium of instruction in any known area. Indeed, it has been suggested that having a school with teachers who can not speak the relevant ethnic language seems to seriously hamper the educational performance of minority children, who largely speak their own indigenous language at home, and may have little or no exposure to Vietnamese before they arrive at primary schools (e.g., Ministry of Education and Training, UNICEF and UNESCO, 2008).

Third, majority households often outperform minority households in gaining from the availability of market and off-farm employment. Our results show that the existence of a daily market does not seem to help the minority, but significantly supports the majority group. This result is broadly consistent with the standard view in qualitative research that the majority group does better than the minority group in a market context and in market places, and that the majority dominates overall in trading systems (World Bank, 2009). Barriers that the minority often face in the market place include lack of market information, language difficulties and calculation skills, and a culture of community reciprocity which prevents many minorities from doing business in an otherwise or strictly entrepreneurial fashion (see World Bank, 2009a; Tran, 2004; Vietnam Academy of Social Sciences, 2009).

Finally, in terms of off-farm employment, living in a commune with a foreign-owned enterprise within 10 km, has the largest positive impact among all infrastructure facilities on the majority, but no impact on the minority. On the other hand, minority households appear to benefit more than majority counterparts from presence of local enterprises. Lack of social networks, poor infrastructure, their isolated nature and limited mobility of many minority households are likely key drivers for this result.

2.5.5 Robustness Checks

Before turning to decomposition of ethnic gap, which relies on estimated results, we subject our estimates a few robustness checks. The approach here is twofold, concentrating on two concerns. First, we check if there is sensitivity in the choice of instrument for a key variable, namely "requires interpretation." Second, we exclude outliers from our sample, in particular, households with remittances (often cash-transfers from overseas relatives) and households who are classified as poor in the survey year, or the year prior.

For the first concern, the Vietnamese speaking ability of a household may be affected by not only the commune-average Vietnamese speaking ability, which we have controlled for using within-commune deviation as an instrument. Rather, it may also depend on how assimilated to Kinh group is that household and/or the specific ethnicity it belongs to. At the household level, the degree of assimilation may come from intermarriage with a Kinh partner (at present or in the past). Unfortunately, we do not have information on ethnicity of any household member rather than the household head, so we are not being able to assess this effect. For a specific ethnicity, we do know that the better off it is, the more assimilated it tends to be to the Kinh majority (Baulch et al., 2007).

We address this concern by constructing an instrument which measures the propensity to speak Vietnamese for each ethnicity,[10] checking to see if our main results are sensitive to this choice of instrument. Results are presented in Table 2.8, column (2). For brevity, only results in pooled and regression differences are presented. As we might expect, we find larger negative impact from a language barrier. Not being able to speaking Vietnamese is now associated with a 20 percent reduction in expenditure as compared to 14 percent using the within-commune deviation as an instrument. This result further underpins the importance of language in the minority living standard. Other variables in our regression remain stable in their contribution to the ethnic gap, with an exception to the distance to the city which now helps narrow the ethnic gap.

For the second concern, as discussed in the background section above, households classified as poor are likely to receive support from various government policies, such as the HEPR program. Furthermore, given remnants of the war in Vietnam as well as the Vietnamese culture of reciprocity, some households may receive overseas and/or domestic remittances. This external support, regardless of whether it originates from relatives or through the government, would substantially alter expenditures and the way a household gains from its endowments.

In Table 2.8, we exclude each group of outliers in turn. In column (3), for the restricted sample of households not receiving remittances, only a negligible ethnic difference is observed in returns to household capital as compared with that estimated for the full sample. In particular, this ethnic difference falls slightly, undoubtedly since many more majority households, rather than minority ones that receive remittances, are removed from the sample.

Taking out households classified as poor in column (4) leaves coefficients relatively stable. While language barriers become insignificant in this subsample of 'better off' households, disparity in returns to education increases, though marginally. Non-poor minority households tend to be able to make more use of their land and local off-farm opportunities in this subsample.

[10]We use information from income data which has a larger sample to construct the propensity to speak Vietnamese for each ethnicity.

TABLE 2.8: Robustness Checks

	(1) Full sample		(2) 'Require Interpretation' instrumented by propensity to speak Vietnamese		(3) Exclude households with remittances		(4) Exclude poor households	
	Pool	Difference	Pool	Difference	Pool	Difference	Pool	Difference
Household Characteristics								
Household size (log)	-0.333***	0.065	-0.333***	0.063	-0.321***	0.081	-0.347***	0.135
	(-12.99)	(0.96)	(-12.97)	(0.91)	(-12.20)	(1.18)	(-12.13)	(1.50)
Aged from 7 to 16	0.158***	-0.148	0.157***	-0.132	0.151***	-0.151	0.118**	-0.221
	(3.21)	(-1.22)	(3.18)	(-108)	(2.97)	(-1.24)	(2.20)	(-1.58)
Males aged over 16 years	0.482***	-0.096	0.481***	-0.051	0.482***	-0.064	0.381***	-0.148
	(8.31)	(-0.65)	(8.27)	(-0.34)	(8.01)	(-0.44)	(6.09)	(-0.83)
Female aged over 16 years	0.424***	-0.083	0.422***	-0.053	0.436***	-0.039	0.373***	-0.177
	(6.69)	(-0.47)	(6.67)	(-0.30)	(6.61)	(-0.21)	(5.43)	(-0.94)
Parent(s) and three or more children	-0.060***	-0.041	-0.060***	-0.037	-0.061***	-0.031	-0.060***	-0.042
	(-3.54)	(-0.98)	(-3.56)	(-0.88)	(-3.52)	(-0.74)	(-3.18)	(-0.81)
Three-generation households	-0.056***	0.006	-0.056***	0.012	-0.069***	-0.006	-0.050**	-0.015
	(-2.68)	(0.11)	(-2.69)	(0.22)	(-3.19)	(-0.11)	(-2.22)	(-0.23)
Other household structures	0.000	0.037	0.001	0.033	0.004	0.043	-0.011	0.070
	(0.02)	(0.62)	(0.03)	(0.56)	(0.16)	(0.71)	(-0.44)	(0.92)
Headed by female	-0.071***	0.017	-0.071***	0.017	-0.068***	0.000	-0.061***	-0.030
	(-4.11)	(0.35)	(-4.11)	(0.35)	(-3.70)	(0.00)	(-3.27)	(-0.52)
Age of household head	0.011***	0.016**	0.011***	0.016**	0.013***	0.016*	0.011***	0.018*
	(3.26)	(1.97)	(3.26)	(2.01)	(3.75)	(1.92)	(3.05)	(1.85)
Age of household head, squared/100	-0.012***	-0.015*	-0.012***	-0.016*	-0.014***	-0.015*	-0.012***	-0.017*
	(-3.98)	(-1.88)	(-3.98)	(-1.95)	(-4.55)	(-1.82)	(-3.54)	(-1.77)
Household Human Capital								
Require interpretation	-0.141**	.	-0.203***	.	-0.139**	.	-0.098	.
	(-2.57)	.	(-3.91)	.	(-2.41)	.	(-1.20)	.

	(1) Full sample		(2) 'Require Interpretation' instrumented by propensity to speak Vietnamese		(3) Exclude households with remittances		(4) Exclude poor households	
	Pool	Difference	Pool	Difference	Pool	Difference	Pool	Difference
Years of schooling	0.054***	0.023***	0.054***	0.023***	0.054***	0.019***	0.049***	0.026***
	(24.06)	(3.60)	(24.06)	(3.68)	(23.54)	(3.00)	(19.80)	(3.41)
Household Physical Capital								
Irrigated crop land	0.009***	0.003	0.009***	0.003	0.008***	0.003	0.007***	0.005*
(1000 m2)	(5.73)	(0.96)	(5.74)	(0.91)	(5.55)	(1.03)	(4.64)	(1.85)
Non-Irrigated crop land	0.006***	-0.004	0.006***	-0.004	0.006***	-0.003	0.004**	-0.002
(1000 m2)	(3.28)	(-1.21)	(3.28)	(-1.25)	(3.30)	(-0.85)	(2.34)	(-0.67)
Perennial land	0.009***	-0.005*	0.009***	-0.005*	0.009***	-0.004	0.008***	-0.007**
(1000 m2)	(7.35)	(-1.75)	(7.37)	(-1.75)	(7.18)	(-1.56)	(6.95)	(-2.02)
Forestry land	0.001***	0.000	0.001***	0.000	0.001**	-0.000	0.001**	0.001
(1000 m2)	(2.84)	(0.12)	(2.84)	(0.13)	(2.10)	(-0.13)	(2.27)	(0.73)
Water-surface land	0.012***	-0.010**	0.012***	-0.010**	0.011***	-0.004	0.011***	-0.011**
(1000 m2)	(4.20)	(-2.17)	(4.20)	(-2.16)	(3.78)	(-0.21)	(4.13)	(-2.26)
Other land	0.006*	0.010	0.006*	0.010	0.006*	0.010	-0.000	-0.079***
(1000 m2)	(1.81)	(0.48)	(1.81)	(0.50)	(1.73)	(0.49)	(-0.06)	(-2.63)
Commune characteristics								
Rural high or low mountains	-0.039**	0.078	-0.038*	0.045	-0.022	0.080	-0.011	0.032
	(-1.98)	(1.10)	(-1.95)	(0.61)	(-1.11)	(1.16)	(-0.54)	(0.37)
Rural hilly lands	0.015	-0.045	0.015	-0.041	0.031	-0.059	0.027	0.042
	(0.51)	(-0.39)	(0.52)	(-0.35)	(1.07)	(-0.54)	(0.88)	(0.23)
Distance to city	-0.287***	-0.112	-0.268***	-0.327**	-0.295***	-0.168	-0.238***	0.006
(1000 km)	(-5.04)	(-0.81)	(-4.84)	(-2.42)	(-5.13)	(-1.22)	(-3.82)	(0.04)
Commune Infrastructure								
Power	0.170***	-0.025	0.172***	-0.034	0.163***	-0.017	0.130*	-0.038
	(2.85)	(-0.16)	(2.85)	(-0.22)	(2.61)	(-0.11)	(1.73)	(-0.18)

	(1) Full sample		(2) 'Require Interpretation' instrumented by propensity to speak Vietnamese		(3) Exclude households with remittances		(4) Exclude poor households	
	Pool	Difference	Pool	Difference	Pool	Difference	Pool	Difference
Clean water	0.031**	-0.096	0.031**	-0.087	0.026*	-0.095	0.026*	-0.090
	(2.06)	(-1.49)	(2.10)	(-1.23)	(1.76)	(-1.54)	(1.66)	(-1.18)
Hard-surfaced car road	0.004	-0.078**	0.004	-0.091**	0.002	-0.076*	0.004	-0.077*
	(0.28)	(-1.99)	(0.30)	(-2.31)	(0.16)	(-1.94)	(0.29)	(-1.72)
Primary school	0.028**	0.041	0.029**	0.019	0.025*	0.050	0.021	0.057
	(2.02)	(1.24)	(2.08)	(0.56)	(1.84)	(1.52)	(1.44)	(1.54)
Village daily market	0.048***	-0.025	0.048***	-0.033	0.049***	-0.015	0.047***	-0.034
	(3.23)	(-0.50)	(3.22)	(-0.64)	(3.28)	(-0.29)	(3.07)	(-0.59)
State-owned enterprise	0.021	-0.033	0.022	-0.023	0.015	-0.035	0.023	-0.034
within 10km	(1.28)	(-0.66)	(1.25)	(-0.63)	(0.90)	(-0.72)	(1.31)	(-0.70)
Foreign-shared enterprise	0.091***	0.090	0.092***	0.109	0.096***	0.085	0.090***	0.082
within 10km	(3.95)	(1.13)	(3.99)	(1.46)	(4.15)	(1.15)	(3.87)	(1.11)
Local enterprise	0.061***	-0.014	0.062***	-0.030	0.060***	-0.018	0.061***	-0.072*
	(4.23)	(-0.37)	(4.27)	(-0.63)	(4.20)	(-0.49)	(4.12)	(-1.95)
Minority Group	-0.242***	.	-0.227***	.	-0.249***	.	-0.233***	.
	(-9.00)	.	(-9.46)	.	(-9.06)	.	(-7.64)	.
Constant	7.281***	-0.304	7.292***	-0.282	7.218***	-0.315	7.495***	-0.408
	(71.81)	(-1.17)	(71.46)	(-1.09)	(69.58)	(-1.24)	(64.76)	(-1.30)
Number of households	6560	6560	6560	6560	6235	6235	5366	5366
Communes	2187	2187	2187	2187	2180	2180	2120	2120

Log of real per capita expenditure as dependent variable. t statistics are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression omits the proportion of members aged 06; one- or two-generation households that consist of no more than two children; commune geographical types as coastal/delta; other dummy variables having yes=1 and no=0.

2.5.6 Decomposition of the Ethnic Gap

In this section, we employ a Blinder-Oaxaca decomposition (Oaxaca, 1973; Blinder, 1973), commonly applied in labour economics, to investigate how household and commune characteristics and endowments contribute to differences between the majority and minority groups. This technique has been applied in previous research on ethnic inequality in Vietnam (Van de Walle and Gunewardena, 2001; Baulch et al., 2007; Hoang et al., 2007; Baulch et al., 2012, 2010). The basic idea is to split (in this case) the average log expenditure gap between the majority and minority populations into the gap caused by the difference in endowments (the regressors) and the difference in returns to endowments (the regression parameters), or

$$\begin{aligned}
\hat{Y}_m - \hat{Y}_e &= (\bar{X}_m\hat{\beta}_m + \bar{Z}_m\hat{\gamma}_m + \hat{\alpha}_m) - (\bar{X}_e\hat{\beta}_e + \bar{Z}_e\hat{\gamma}_e + \hat{\alpha}_e) \\
\hat{Y}_m - \hat{Y}_e &= (\bar{X}_m - \bar{X}_e)\hat{\beta}_m + \bar{X}_e(\hat{\beta}_m - \hat{\beta}_e) \\
&\quad +(\bar{Z}_m - \bar{Z}_e)\hat{\gamma}_m + \bar{Z}_e(\hat{\gamma}_m - \hat{\gamma}_e) + (\hat{\alpha}_m - \hat{\alpha}_e) \qquad (2.5) \\
\hat{Y}_m - \hat{Y}_e &= (\bar{X}_m - \bar{X}_e)\hat{\beta}_m + (\bar{Z}_m - \bar{Z}_e)\hat{\gamma}_m \\
&\quad +\bar{X}_e(\hat{\beta}_m - \hat{\beta}_e) + \bar{Z}_e(\hat{\gamma}_m - \hat{\gamma}_e) + (\hat{\alpha}m - \hat{\alpha}_e) \qquad (2.6)
\end{aligned}$$

where the subscripts m and e stand for majority and minority groups, respectively.

The first two terms on the right-hand side of 2.5 represent the gap in household expenditures that can be attributed to differences in the household-level explanatory variables of the majority and minority groups, and the returns to given household-level explanatory variables. The third and the forth terms measure the gap in expenditures that can be attributed to differences in the commune-level explanatory variables of the two groups, and the returns to those commune-level explanatory variables. The last term, the difference in the constants, represents the difference in unobserved commune fixed effects combined with joint effects of excluded dummy variables.

Rearranging equation 2.5 now gives the left-hand side as the total difference in expenditures, or the ethnic gap, while the sum of the first two terms on the right-hand side of equation 2.6 is the difference in endowments, and the remainder of the expression is the difference in returns to endowments. The results presented in Table 2.9 are arranged in accordance with estimation results presented in Table 2.4, which include household characteristics, household human capital, household physical capital, commune characteristics and commune infrastructure. We use, alternatively, the majority and minority coefficients as reference weights. We also present our decomposition results in comparison with those generated by OLS and FE estimates, which were previously applied in the literature.

Our results indicate that the ethnic gap is largely explained by differences in characteristics and endowments. Together, they account for as much as 66 percent of the ethnic gap using minority group coefficients as the reference weight. When majority coefficients are used as the reference, language was omitted because its coefficient is undefined for the majority. In this case, differences in endowments account for about 49 percent of the ethnic gap.

Regardless of which group is used as a reference, our IV-based decomposition results are in sharp contrast to FE-based decomposition results in Table 2.9 as well as established views in the literature. In particular, the FE-based decomposition results in Table 2.9 suggest that differences in characteristics and endowments explain only about 30 percent of the ethnic gap. These results are similar to findings in the literature using FE estimators on earlier household data (Hoang et al., 2007; Baulch et al., 2007; Van de Walle and Gunewardena, 2001), and again, they are much different compared to our IV-based decomposition results. The reason for the difference between IV-based and FE-based decomposition results is that the FE estimator fails to estimate impacts of commune-specific observed attributes such as geography and infrastructure, and hence misses an important part of the story on inequality.

OLS-based decomposition results, on the other hand, are closer to our IV-based decomposition ones because they can also estimate returns to commune-specific attributes. However, the estimation bias of OLS causes a corresponding bias in the OLS-based decomposition results, which is especially clear in differences in returns to endowments. For example, IV-based (and FE-based) decomposition results suggest the principal driver in the difference in returns to endowments is the ethnic disparity in returns to human capital. To put it differently, according to these two consistent estimators, even if the minority had the same level of human capital (i.e., measured by education and language), they could never 'catch up' with the majority since the minority has lower returns to those attributes. On the contrary, OLS-based decomposition results show an insignificant and much smaller difference in returns to human capital between the two groups. The biased OLS-based decomposition results reported in Table 2.9 are also comparable to OLS-based decomposition findings in previous literature (Baulch et al., 2012, 2010).

It is important to note that not only does the difference in endowments in our IV-based decomposition explain most of the ethnic gap, all of its components are also highly significant. Indeed, with an exception of land, all differences in endowments contribute to widening the ethnic gap. In particular, the differences in infrastructure and household characteristics, each accounting for 20 percent of the ethnic gap, are the largest among all of the components. Differences in household human capital also play an important role in widening the ethnic gap, explaining about 16 percent of the total difference, as do differences in commune characteristics. In summary, our results support government investment in commune infrastructure to mitigate ethnic inequality and add a further call for removing language barriers, improving education for the minority in both the number of years and (presumably) the quality of education as well as allowing for more mobility, especially among the minority.

On the other hand, the components in differences in returns to endowments in our IV-based decomposition, although they contribute to narrowing the ethnic gap, are largely insignificant. A key exception is household human capital, which stands out as highly significant and widens the gap by as much as 40 percent. Returns to household characteristics also increase the ethnic gap, but again with no statistical significance. Finally, the effects of location can also be observed in differences in returns to commune characteristics as well as in the constants, in addition to differences in commune characteristics. Both differences in returns to commune characteristics as well as in the constants are insignificant and apparently do not contribute to the gap.

TABLE 2.9: Decomposition of Ethnic Gap

	Minority Group as Reference			Majority Group as Reference		
	OLS	FE	IV	OLS	FE	IV
Difference in Endowments	0.428***	0.187***	0.375***	0.293***	0.173***	0.276***
	(0.034)	(0.026)	(0.035)	(0.022)	(0.017)	(0.025)
Household Characteristics	0.125***	0.123***	0.119***	0.104***	0.112***	0.112***
	(0.014)	(0.013)	(0.015)	(0.008)	(0.008)	(0.009)
Household Human Capital	0.121***	0.104***	0.089***	0.088***	0.091***	0.090***
	(0.014)	(0.020)	(0.020)	(0.009)	(0.008)	(0.009)
Household Physical Capital	-0.036***	-0.039***	-0.043***	-0.025*	-0.030**	-0.029**
	(0.009)	(0.011)	(0.011)	(0.013)	(0.013)	(0.014)
Commune Characteristics	0.093**	.	0.093**	0.048***	.	0.048***
	(0.037)	.	(0.048)	(0.015)	.	(0.015)
Commune Infrastructure	0.124***	.	0.118***	0.079***	.	0.054***
	(0.024)	.	(0.032)	(0.010)	.	(0.012)
Difference in Returns to Endowments	0.137***	0.377***	0.190***	0.272***	0.392***	0.289***
	(0.034)	(0.026)	(0.034)	(0.023)	(0.018)	(0.027)
Household Characteristics	0.375**	0.348*	0.386	0.397**	0.359**	0.394*
	(0.177)	(0.183)	(0.205)	(0.176)	(0.183)	(0.204)
Household Human Capital	0.048	0.198***	0.223***	0.082*	0.211***	0.221***
	(0.048)	(0.051)	(0.061)	(0.043)	(0.047)	(0.052)
Household Physical Capital	0.007	0.001	-0.005	-0.003	-0.008	-0.019
	(0.011)	(0.009)	(0.010)	(0.021)	(0.020)	(0.022)
Commune Characteristics	-0.020	.	-0.004	0.026	.	0.041
	(0.024)	.	(0.029)	(0.059)	.	(0.075)
Commune Infrastructure	0.042	.	-0.094	0.087	.	-0.030
	(0.078)	.	(0.165)	(0.069)	.	(0.150)
Constant	-0.317	-0.170	-0.317	-0.317	-0.170	-0.317
	(0.195)	(0.172)	(0.255)	(0.195)	(0.172)	(0.254)

Standard errors are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.6 Conclusion

This paper has examined what drives the gap in household expenditures between the two ethnic groups in Vietnam. Based on an IV approach, we have argued that removing language barriers can significantly narrow the gap, that majority households do not benefit more than minorities from local infrastructure investment, and that the ethnic gap is as much as 49 to 66 percent attributed to differences in characteristics and endowments, especially household characteristics and commune infrastructure. The results represent a significant improvement over OLS estimates in explaining the sources of ethnic inequality, as well as FE estimates, which are unable to estimate commune specific observed explanatory variables.

Minorities tend to live in more remote and mountainous areas of Vietnam but this fact is found to contribute a much smaller effect than expected to the ethnic gap. That said, a relaxed policy on migration, especially for the minority would help ease this disparity. On the other hand, minority households, regardless of location, can benefit greatly from basic infrastructure in the commune: clean water, electric power, and the provision of a hard-surfaced road. Government programs, as such Program 135, will thus contribute to enhancing the living standards of minorities and reduce the ethnic gap through providing more equalised levels of commune endowments.

We have shown that while removing language barriers eases ethnic inequality, especially through the gains from schooling and to a lesser extent, from land, the ethnic difference in returns to education are intensified in an environment where households from both groups are found to live. Our results thus suggest that either the special educational needs of minority children are not being addressed in classrooms, or unequal treatment in favour of the majority exists in the labour market.

The importance of language is a new and perhaps surprising addition to the understanding of the differences in expenditures among majority and minority households in Vietnam. Much of this is potentially attributable to the quality of 2006 VHLSS data, and the use of a language indicator in our estimates, but the size of the effect both in the estimation and in the Blinder-Oaxaca decomposition indicates its relative significance, and greatly conditions how we might interpret results from estimates using earlier versions of VHLSS data.

With this mind, there are at least two issues worth exploring in future work. First, it would be useful to further investigate the effects of social characteristics on household living standards. Community and social effects, along with language skills and the ability to 'social network', are important indicators of differences in living standards and potential changes in the poverty rate. Little in our current work accounts for network and community effects, except for the simple presence of language barriers and the co-location that goes with being in a mixed commune. Second, there is a need to investigate the role of labour mobility and barriers in the labour market that prevent more equalised returns from education. These are partly bound up in language differences, but in many circumstances this may not be the case. In this regard, our results would also benefit from measures of the differences in labour market experience among households, as well as a further investigation of potential preferential treatment for the majority (in particular) in mixed communes.

Chapter 3

The Fixed-Effects Vector Decomposition: A critical analysis

3.1 Introduction

Researchers in many fields seek to exploit the advantages of panel data. Having repeated observations across time for each group in a panel allows one, under suitable assumptions, to control for unobserved heterogeneity across the groups which might otherwise bias the estimates. However, traditional panel analysis techniques have difficulty when some of the explanatory variables have little or no variation across time within a group. We consider here the properties of a recently introduced methodology for panel data, known as fixed-effects vector decomposition FEVD, which Plümper and Troeger (2007a) developed to produce improved estimates for such time-invariant or slowly-changing variables.

From the earliest days of panel data methods, there has been a tension between alternative treatments of the unobserved individual effects as either fixed or random. Mundlak (1978) demonstrated that a generalized least squares approach to unobserved group effects, which treats them as random and potentially correlated with the regressor, gives rise to the traditional fixed-effects (FE) estimator. However, FE is a blunt instrument for controlling for correlation between observed and unobserved characteristics because it ignores any systematic average differences between groups. Thus any potential explanatory factors that are constant longitudinally (time-invariant) will be ignored by the FE estimator. Likewise, any explanatory variables that have little within variation (that is, slowly-changing over the longitudinal dimension) will have little explanatory power, and will result in imprecise coefficient estimates that have large standard errors.

Hausman and Taylor (1981) had previously shown that a better estimator than FE is available if some of the explanatory variables are known to be uncorrelated with the unobserved group effect, thus described as *exogenous* explanatory variables. The Hausman-Taylor (HT) estimator is an instrumental variables (IV) procedure that combines aspects of both fixed-effects and random-effects estimation. Given a sufficient number of exogenous regressors, the HT procedure allows time-invariant variables to be kept in the model. It also provides more efficient estimates than FE for the coefficients of the exogenous time-varying variables. The downside of the HT estimator resides in specifying the exogeneity status for each of the time-varying and time-invariant variables in the model. In many practical applications such detailed specification is onerous.

Plümper and Troeger introduced FEVD as an alternative that seemed to be superior to HT because it requires fewer explicit assumptions yet seemed to always have more desirable sampling properties. Like the FE estimator, and unlike HT, FEVD does not require specifying the exogeneity status of the explanatory variables. Like the HT estimator, and unlike FE, the FEVD procedure gives coefficient estimates for time-invariant (and slowly-changing) variables as well as the time-varying variables. Plümper and Troeger motivated the FEVD procedure on heuristic grounds, and advocated it on the strength of favorable results in a Monte Carlo simulation study. In particular, the simulation indicated that FEVD has superior sampling properties for time-invariant explanatory variables.

Although the FEVD procedure comes out of the empirical political science literature, it is rapidly finding application in many other areas including social research and economics. At last count there were well over 400 references in Google Scholar to this emerging estimation methodology. Several empirical studies report standard errors for FEVD-based estimates that are strikingly smaller than estimates based on traditional methods. There is, however, little formal analysis of the FEVD procedure in this literature.

This chapter is a remedy to the lack of formal analysis. We demonstrate that the FEVD coefficient estimator can be equivalently written as an IV estimator, which serves to demystify the nature of the three-stage FEVD procedure and its relationship with other estimators. As one immediate benefit, the IV representation allows us to draw on a standard toolkit of results.

First, using the IV variance formula, we show that the FEVD standard errors for coefficients of both the time-varying and time-invariant variables are uniformly *too small.* In the case of the latter variables, the discrepancy in the FEVD standard errors is *unbounded*, and grows with the length of the panel and with the variance of the group effects.

Second, using the moment-condition representation, we prove that the coefficients of the time-varying variables in FEVD are exactly the same as in FE. This result is apparent in many of the practical studies which list FE estimates beside FEVD estimates, but it is hardly mentioned in the existing analytical material. An immediate implication is that FEVD estimates, like FE, are *inefficient* if any of the time-varying variables are exogenous.

Third, FEVD usually produces lower variance estimates of time-varying coefficients than HT in small samples. However, it does so by including invalid instruments that produce *inconsistent* estimates. So, even with massive quantities of data those FEVD estimates will deviate from the truth.

Further developments can also be made to the estimator, to exploit the ideas in FEVD while avoiding the problems of that procedure. The advantage of FEVD will be found in smaller samples where the large sample concept of consistency does not dominate. The Monte Carlo simulation studies by Plümper and Troeger (2007a) and Mitze (2009) show a trade-off between bias and efficiency in which FEVD often appears to be better than either FE or HT under quadratic loss. We present Monte Carlo evidence that a standard shrinkage approach combines the desirable small sample properties of FEVD with the desirable large sample properties of the HT estimator, so that it has superior risk to both FEVD and HT over a wide region of the parameter space.

In the next section we introduce the notation to be used and describe the three-stage FEVD estimator. We summarize the connections between these stages in a theorem, which we prove by comparing the various moment conditions. This approach demonstrates naturally the description of the FEVD estimator as IV. Section 3.3 compares the correct IV variance formula with the formula implicit in the standard errors of the three-stage FEVD approach. The main results are summarized in two further theorems. We also provide an empirical example and Monte Carlo simulations to illustrate these results and some from the previous section. In Section 3.4 we examine the relationships between estimators in more detail, allowing the possibility of a trade-off between bias and variance to produce an estimator with lower mean-squared error. Some Monte Carlo evidence in the spirit of Plümper and Troeger that demonstrates the superiority of a standard shrinkage estimator is also provided in this section. Section 3.5 has some overall conclusions.

3.2 The Model

The data are ordered so that there are N groups each of T observations. The model for a single scalar observation is

$$y_{it} = X_{it}\beta + Z_i\gamma + u_i + \epsilon_{it} \text{ for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T. \tag{3.1}$$

Here, X_{it} is a $k \times 1$ vector of time-varying explanatory variables, and Z_i is a $p \times 1$ vector of time-invariant explanatory variables.[1] The parameters β, γ, the group effect u_i, and the error term ϵ_{it} are all unobserved. Some elements of X_{it} or Z_i are correlated with the group effect u_i, in which case we call those variables *endogenous.* Otherwise we call those variables *exogenous.* With endogenous explanatory variables standard linear regression techniques may produce estimates of the unknown parameters which are inconsistent in the sense that they do not converge to the true parameter values as the sample size grows large. One standard approach to this endogeneity problem is to use the instrumental variables technique developed by Hausman and Taylor.

Notation

The presentation is considerably simplified by introducing some projection matrix notation. Let

$$D = I_N \otimes \iota_T, \tag{3.2}$$

where I_N is an $N \times N$ identity matrix and ι_T is a $T \times 1$ vector of ones. That is, D is a matrix of dummy variables indicating group membership. For any matrix M, we use $P_M = M(M'M)^{-1}M'$ to indicate the projection matrix for M, and we use $Q_M = I - P_M$ to indicate the projection matrix for the nullspace of M. For example,

$$P_D = D(D'D)^{-1}D' = \frac{1}{T}(I_N \otimes \iota_T\iota_T') \tag{3.3}$$

[1] The setup here describes a balanced panel with observations on every t for each i, but the ideas extend to unbalanced panels with more complicated notation. A constant can be represented in this model by including a vector of ones as part of the time-invariant elements, Z.

is the matrix which projects a vector onto D. This particular projection produces a vector of group means. That is, $P_D y = \{\bar{y}_i\} \otimes \iota_T$, where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$. Also,

$$Q_D = I_{NT} - P_D \tag{3.4}$$

is the matrix which produces the within-group variation. That is, $Q_D y = \{y_{it} - \bar{y}_i\}$ is the $NT \times 1$ vector of within-group differences.

The FEVD Estimator

The FEVD proceeds in three stages, which we detail below. To sharpen the analysis, we assume that the elements of Z are exactly time-invariant (not just slowly-changing), so that $P_D Z = Z$. An explicit analysis of the slowly-changing case yields qualitatively similar insights.

Stage 1

Perform a fixed effects regression of y on the time-varying X. This is ordinary least squares (OLS) after within transformation:

$$Q_D y = Q_D X\beta + Q_D Z\gamma + Q_D u + Q_D e \tag{3.5}$$

Note that the within transformation wipes out the time-invariant explanatory variables Z, and the group effect u. The moment condition corresponding to a fixed effects regression is

$$(y - Xb)' Q_D X = 0. \tag{3.6}$$

The unexplained component after this first step, the residuals from FE estimation, is $y - Xb$. The group-average of the unexplained component is $P_D(y - Xb)$.

Stage 2

Regress the group-average of the unexplained component from the first step on the time-invariant Z. The moment condition is $\big(P_D(y - Xb) - Zg\big)' Z = 0$. Using the fact that $P_D Z = Z$, this moment condition can be equivalently written as

$$(y - Xb - Zg)' Z = 0. \tag{3.7}$$

The group-average residuals from this regression are

$$h = P_D(y - Xb - Zg). \tag{3.8}$$

Stage 3

The group-average residuals h are used as proxies for the unobserved effect u in the original model. Regress y on X, Z, and h to get β, γ, and δ where δ is the coefficient of h. The coefficients from this step are the final FEVD estimates. The moment conditions are

$$(y - X\beta - Z\gamma - h\delta)'[X, Z, h] = 0. \tag{3.9}$$

Theorem 1. *The solution for β is b from Stage 1; the solution for γ is g from Stage 2; and the solution for δ is one.*

Proof. We need to verify that the moment conditions (3.9) are satisfied at $\beta = b$, $\gamma = g$, and $\delta = 1$. This requires that

$$(y - Xb - Zg - h)'[X, Z, h] = 0. \tag{3.10}$$

Substituting in the definition of h from (3.8) and gathering terms, this simplifies to

$$(y - Xb - Zg)'Q_D[X, Z, h] = 0. \tag{3.11}$$

Using the fact that $Q_DZ = 0$, this further simplifies to

$$(y - Xb)'Q_D[X, Z, h] = 0. \tag{3.12}$$

The first set of equalities in (3.12) must be satisfied, since it is identical to the moment condition (3.6) that defines b. The second set of equalities must be satisfied since $Q_DZ = 0$. Similarly, the third set of equalities must be satisfied since $Q_Dh = 0$, which follows from the definition of h in (3.8) and the fact that $Q_DP_D = 0$. □

Instrumental Variables Representation

Using Theorem 1 we can show that the FEVD estimator can also be expressed as an IV estimator for a particular set of instruments. The major benefit of using the

IV representation is that one can draw on a standard toolkit of results. Theorem 1 shows that the FEVD estimates of β are identical to the standard fixed effects estimator b from Stage 1. This estimator is defined by the moment condition (3.6). Theorem 1 also shows that the FEVD estimates of γ are equivalent to the estimator of g from Stage 2. This estimator is defined by the moment condition (3.7). Combining both moment conditions, and using the fact that $Q_D Z = 0$, the full moment conditions for the FEVD estimator are

$$(y - X\beta - Z\gamma)'[Q_D X, Z] = 0. \tag{3.13}$$

In other words, the FEVD estimator is equivalent to an IV estimator using the instruments $Q_D X$ and Z.

3.3 Variance Formulae

3.3.1 IV Sampling Variance

Using standard results for IV estimators, the asymptotically correct sampling variance of the FEVD procedure is

$$V_{\text{IV}}(\beta, \gamma) = (H'W)^{-1} H'\Omega H (W'H)^{-1} \text{ for } H = [Q_D X, Z] \text{ and } W = [X, Z]. \tag{3.14}$$

Here, H is the matrix of instruments and W is the matrix of explanatory variables. Ω is the covariance of the residual, $u_i + \epsilon_{it}$, which can be expressed as

$$\Omega = \sigma_\epsilon^2 I_{NT} + \sigma_u^2 I_N \otimes \iota_T \iota_T' = \sigma_\epsilon^2 Q_D + (\sigma_\epsilon^2 + T\sigma_u^2) P_D. \tag{3.15}$$

Using straightforward algebraic manipulation of (3.14), the *IV* sampling variances of β and of γ are expanded as below:

$$(H'W) = \begin{pmatrix} X'Q_D \\ Z' \end{pmatrix} [X, Z] = \begin{pmatrix} X'Q_D X & 0 \\ Z'X & Z'Z \end{pmatrix}$$

$$(H'W)^{-1} = \begin{pmatrix} (X'Q_D X)^{-1} & 0 \\ -(Z'Z)^{-1} Z'X (X'Q_D X)^{-1} & (Z'Z)^{-1} \end{pmatrix}$$

$$H'\Omega H = \begin{pmatrix} X'Q_D \\ Z' \end{pmatrix} [\sigma_\epsilon^2 Q_D + (\sigma_\epsilon^2 + T\sigma_u^2)P_D][Q_D X, Z] = \begin{pmatrix} \sigma_\epsilon^2 X'Q_D X & 0 \\ 0 & (\sigma_\epsilon^2 + T\sigma_u^2)Z'Z \end{pmatrix}$$

$$(H'W)^{-1}H'\Omega H = \begin{pmatrix} \sigma_\epsilon^2 I & 0 \\ -\sigma_\epsilon^2 (Z'Z)^{-1}Z'X & (\sigma_\epsilon^2 + T\sigma_u^2)I \end{pmatrix}$$

$$V_{\text{IV}}(\beta,\gamma) = (H'W)^{-1}H'\Omega H(W'H)^{-1} =$$

$$\begin{pmatrix} \sigma_\epsilon^2 (X'Q_D X)^{-1} & -\sigma_\epsilon^2 (X'Q_D X)^{-1} X'Z(Z'Z)^{-1} \\ -\sigma_\epsilon^2 (Z'Z)^{-1}Z'X(X'Q_D X)^{-1} & \begin{matrix}(\sigma_\epsilon^2 + T\sigma_u^2)(Z'Z)^{-1} \\ +\sigma_\epsilon^2 (Z'Z)^{-1}Z'X(X'Q_D X)^{-1}X'Z(Z'Z)^{-1}\end{matrix} \end{pmatrix} \tag{3.16}$$

3.3.2 FEVD Sampling Variance in Comparison with IV Sampling Variance

Plümper and Troeger state that the sampling variance of the FEVD estimator can be obtained by applying the standard OLS formula to the Stage 3 regression. Therefore,

$$V_{\text{FEVD}}(\beta,\gamma,\delta) = s^2([X,Z,h]'[X,Z,h])^{-1} = s^2 \begin{pmatrix} X'X & X'Z & X'h \\ Z'X & Z'Z & Z'h \\ h'X & h'Z & h'h \end{pmatrix}^{-1} . \tag{3.17}$$

Here, $s^2 = \|y - X\beta - Z\gamma - h\|^2/dof$, where dof is the degrees of freedom. By application of (3.8), the expression for s^2 can be simplified to

$$s^2 = \|Q_D(y - X\beta)\|^2/dof, \tag{3.18}$$

which we note is the standard textbook FE estimator for σ_e^2 when $dof = NT-N-k$ (see *e.g.* Wooldridge, 2002, p. 271).[2]

Now consider the variance of β. The FEVD variance formula for β is the top-left block of the overall FEVD variance formula in (3.17); using the partitioned-inverse formula this submatrix can be written as

$$V_{\text{FEVD}}(\beta) = s^2(X'Q_{[Z,h]}X)^{-1}. \tag{3.19}$$

From the IV variance formula in (3.16), the correct variance for β is

$$V_{\text{IV}}(\beta) = \sigma_e^2(X'Q_D X)^{-1}. \tag{3.20}$$

Note that this is exactly the textbook fixed effects variance formula.

Now we note from (3.18) that s^2 is a consistent estimator of σ_e^2. However, the matrices in the FEVD formula (3.19) and the correct IV formula (3.20) differ. The FEVD variance formula for β must therefore be incorrect, and we can show the direction of the error.

Theorem 2. *The FEVD variance formula for coefficients on time-varying variables is too small.*

Proof. Now $P_D[Z,h] = [Z,h]$, so that $P_D P_{[Z,h]} = P_{[Z,h]}$. Such a relationship between projection matrices implies that $P_D - P_{[Z,h]}$ is positive semi-definite (in matrix shorthand, $P_D \geq P_{[Z,h]}$). So, $Q_D \leq Q_{[Z,h]}$. That $(X'Q_{[Z,h]}X)^{-1} \leq (X'Q_D X)^{-1}$ follows immediately. This inequality will almost always be strict because the $p+1$ variables $[Z,h]$ cannot span the whole of the N-dimensional space of group operator D, and the X's have arbitrary within-group variation. □

The FEVD formula for the variance of β is *biased* in that it systematically *understates* the true sampling variance of the estimator. The essential inequality does not disappear as N gets larger, so the formula is also *inconsistent*. The usual reported standard errors will be *too small*.

[2] The usual OLS formula for the standard errors from the Stage 3 regression would calculate the scale term using $dof = NT-k-p-1$, where p is the number of Z variables including the constant and the final minus one allows for the additional regressor h. This divisor would clearly produce an inconsistent estimator of σ_e^2 for large N and small T. Plümper and Troeger (2007a, p. 129) mention briefly an adjustment to the degrees of freedom and, although they do not give an explicit formula, their software employs the divisor $dof = NT-N-k-p+1$ (Plümper and Troeger, 2007b). This adjustment would yield a consistent estimate of σ_e^2, but it is nonstandard and slightly biased. To sharpen the subsequent analysis, we use the standard unbiased estimator of σ_e^2, in which $dof = NT-N-k$.

Now, consider the variance of γ. The FEVD variance formula for γ is the middle block of the overall FEVD variance formula in (3.17).

We have

$$V_{\text{FEVD}}(\gamma) = s^2(Z'Q_{[X,h]}Z)^{-1}$$
$$= s^2(Z'Z)^{-1} + s^2(Z'Z)^{-1}Z'[X,h]([X,h]'Q_Z[X,h])^{-1}[X,h]'Z(Z'Z)^{-1}$$

where the second expression comes from the alternative representation of the partitioned inverse. Note that $Z'h = 0$, so in the partitioned central matrix of the second term, only the sub matrix corresponding to the X will be selected. We then have the simplification:

$$V_{\text{FEVD}}(\gamma) = s^2(Z'Z)^{-1} + s^2(Z'Z)^{-1}[(Z'X),0]([X,h]'Q_Z[X,h])^{-1}[(X'Z),0](Z'Z)^{-1}$$

Since $Q_Zh = h$, the right hand side is further simplified to:

$$V_{\text{FEVD}}(\gamma) = s^2(Z'Z)^{-1}+s^2(Z'Z)^{-1}[(Z'X),0]([(Q_ZX)h]'[(Q_ZX)h])^{-1}[(X'Z),0](Z'Z)^{-1}$$

$$= s^2(Z'Z)^{-1} + s^2(Z'Z)^{-1}(Z'X)(X'Q_ZQ_hQ_ZX)^{-1}(X'Z)(Z'Z)^{-1}$$

As $Q_ZQ_hQ_Z = Q_ZQ_h = I - P_Z - P_h$, and as h and Z are orthogonal, $P_Z + P_h = P_{[Zh]}$, the FEVD variance formula in (3.17) written as:

$$V_{\text{FEVD}}(\gamma) = s^2(Z'Z)^{-1} + s^2(Z'Z)^{-1}Z'X\left(X'Q_{[Z,h]}X\right)^{-1}X'Z(Z'Z)^{-1}. \tag{3.21}$$

In contrast, according to (3.14), the correct variance for γ is

$$V_{\text{IV}}(\gamma) = \sigma_e^2(Z'Z)^{-1} + T\sigma_u^2(Z'Z)^{-1} + \sigma_e^2(Z'Z)^{-1}Z'X(X'Q_DX)^{-1}X'Z(Z'Z)^{-1}. \tag{3.22}$$

Again, s^2 is a consistent estimator of σ_e^2, so the first term in (3.21) and in (3.22) is essentially the same. However, the expressions are otherwise different, so the FEVD variance formula for γ must also be incorrect. Again, we can show the direction of the error.

Theorem 3. *The FEVD variance formula for time-invariant variables is too small.*

Proof. As shown in the proof of Theorem 2, $(X'Q_DX)^{-1} \geq (X'Q_{[Z,h]}X)^{-1}$ with almost certain strict inequality, so the last term in the FEVD variance formula (3.21) understates the corresponding term in the correct variance expression (3.22). The only exception would be the unlikely event that X and Z are exactly orthogonal, causing those terms to vanish. But even then, the FEVD variance formula will be an understatement because it omits the term $T\sigma_u^2(Z'Z)^{-1}$, which must be positive definite whenever there are random group effects. □

In general the FEVD variance formula for γ is systematically *biased* and *inconsistent*. The usual reported standard errors will be *too small*. The extent of the downward bias is unbounded. The correct variance expression includes a term that is directly proportional to the number of observations per group T and to the variance of the group effects σ_u^2. In contrast the FEVD variance formula, and hence the standard errors, are unaffected by these parameters. By increasing either or both of these parameters, with everything else held constant, the extent of the downward bias in the FEVD variance formula becomes arbitrarily large.

3.3.3 Empirical Evidence for the Theorems

Reported results from the applied empirical literature align with these theoretical results. Table 3.1 presents our replication of Table 1 in Belke and Spies (2008), and shows results for pooled OLS (OLS), FE, FEVD, and HT. We add a column for the results from Stage 2 of FEVD and a row for the coefficient δ that arises in Stage 3 to further illustrate our theoretical results.[3] The first six variables only are shown for brevity. They include logged nominal GDP of the importing country *lngdim*, logged nominal GDP of the exporting country *lngdpex* and logged bilateral real exchange rate *lrer*, as time-varying variables. The time-invariant variables shown are logged great circle distance in km *ldist*, border length in km *border*, and dummy for one or both countries being landlocked *ll*. Results are estimated from a panel sample of $N = 420$ trading pairs for $T = 14$ years giving 5262 observations.

The coefficients for the first three (time-varying) variables are the same for FE and FEVD, as shown by Theorem 1. To illustrate the second aspect of Theorem 1, the coefficients for the next three (time-invariant) variables are exactly equal in Stage 2 and FEVD, and the solution for δ is one. Theorem 2 is illustrated by the way the first three FEVD reported standard errors are systematically smaller than

[3] We are grateful to those authors for supplying their data. We found some occasional small differences in reported standard errors, probably due to use of "robust" standard errors in the published results. The full table of our replication results is provided in the Appendix 3.6.2.

TABLE 3.1: Partial Replication of Belke and Spies (2008)

	(1) POLS	(2) FE	(3) Stage 2	(4) FEVD	(5) HT
lngdpim	0.88***	0.68***	.	0.68***	0.68***
	(0.04)	(0.11)	.	(0.01)	(0.11)
lngdpex	0.89***	0.71***	.	0.71***	0.71***
	(0.03)	(0.07)	.	(0.01)	(0.07)
lrer	-0.01	0.13**	.	0.13***	0.13**
	(0.01)	(0.06)	.	(0.01)	(0.06)
ldist	-1.27***	.	-1.41***	-1.41***	-1.75***
	(0.11)	.	(0.04)	(0.00)	(0.16)
border	-0.00	.	0.00**	0.00***	-0.00
	(0.00)	.	(0.00)	(0.00)	(0.00)
ll	-0.16*	.	-0.23***	-0.23***	-0.16
	(0.10)	.	(0.04)	(0.00)	(0.14)
	...	...	...	...	...
δ	.	.	.	1.00***	.
	.	.	.	(0.00)	.

One, two, and three asterisks reflect significance at the 0.10, 0.05, and 0.01 confidence levels, respectively. Robust standard errors are in parentheses.

the FE ones, in an order of 0.01, 0.01 and 0.01, against 0.11, 0.07 and 0.06, even though the coefficients themselves are identical and the standard error formula for FE is well established as being correct under the assumptions of the model.

It is a little harder to illustrate Theorem 3, which says that the FEVD standard errors on the time-invariant variables are similarly understated. However the HT estimator is just-indentified in this case, which is the reason the HT coefficients and standard errors for time-varying variables are exactly the same as FE. It is no surprise, then, that the coefficient estimates of three time-invariant variables (which are all exogenous) are generally similar for POLS, FEVD, and HT. As expected, the HT standard errors are slightly larger but very close to those for POLS, in an order of 0.16, 0.00 and 0.14 against 0.11, 0.00 and 0.10. However the FEVD standard errors are very small, at 0.00 in every case for the precision that is shown. This is most implausible, because one would not expect POLS to be generally less efficient, given the structure of this example.

Belke and Spies (2008) is the only paper to our knowledge that reports results for all methods including POLS, FE, FEVD, and HT. Several other applications report both FE and FEVD results (*e.g.* Caporale et al., 2009; Mitze, 2009; Krogstrup and Wälti, 2008). In the studies we examined, the FE t-statistics were consistently smaller than those reported for FEVD time-varying variables — and often much smaller — except for few cases affected by robust standard error formulae. Again, this is despite the fact that the coefficient estimators were actually identical by construction.

3.3.4 Monte Carlo Evidence for the Theorems

We further illustrate theoretical results by simulation examples. We compare the FEVD estimator with an IV estimator having moment conditions specified in (3.13) as well as a correct sampling variance formula specified in (3.16).

The data generating process for our simulation is

$$y_{it} = 1 + 0.5x_1 + 2x_2 - 2.5z_1 + 1.8z_2 + u_i + e_{it}. \tag{3.23}$$

Here, $[x_1, x_2]$ is a time-varying mean-zero orthonormal design matrix, and $[z_1, z_2]$ is a time-invariant mean-zero orthonormal design matrix, all fixed across all experiments. The idiosyncratic error term e is standard normal. The group effect u is drawn from a normal distribution in each replication. All variables are uncorrelated with u, and the variance of u conditional on all variables is 1[4]. Therefore, all variables are treated as known exogenous in our simulation. This assumption will be relaxed in the next section to allow for correlation between some of time-invariant variables with group effect u to further investigate performance of FEVD to alternative estimators.

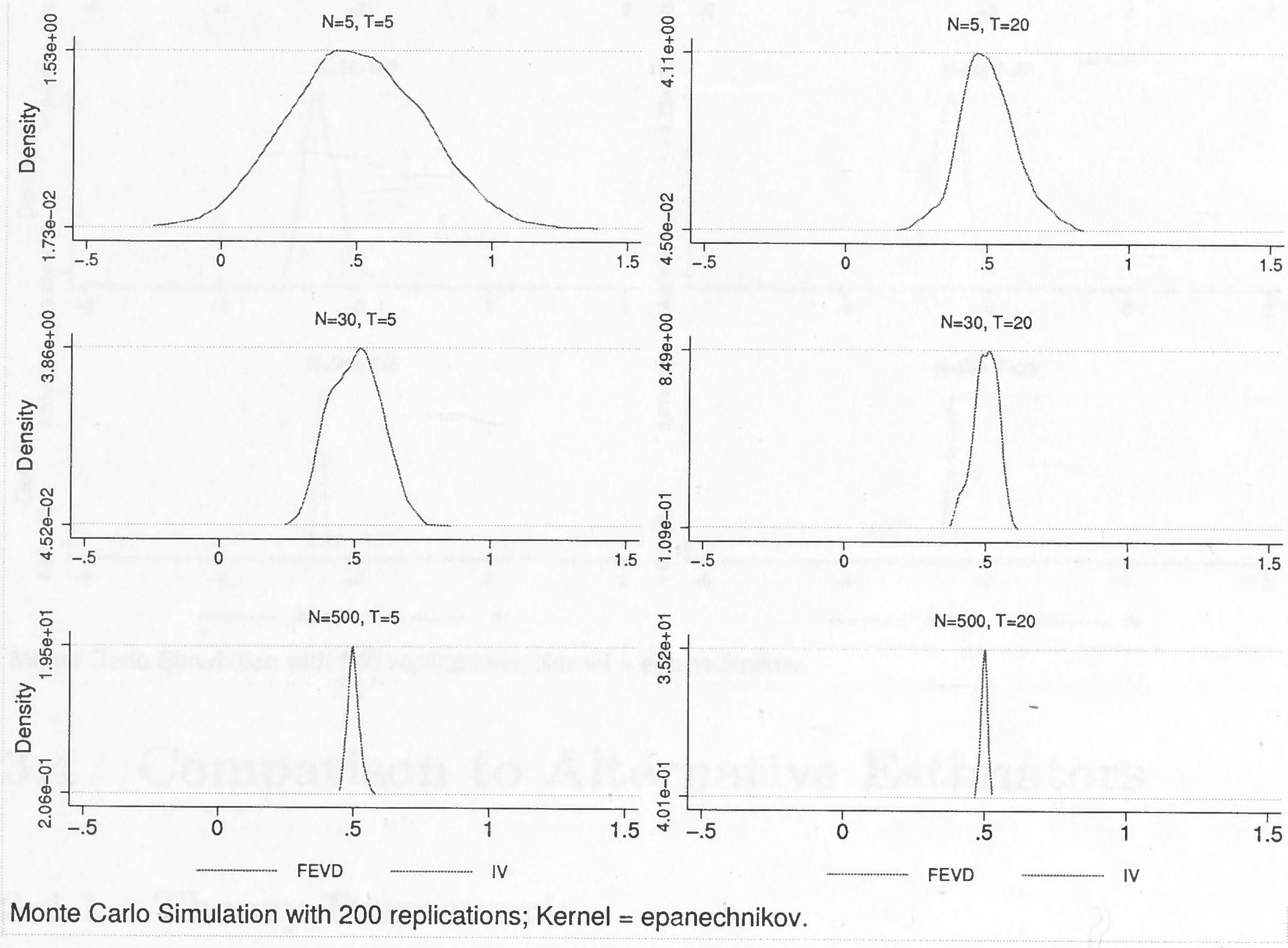
Each experiment in our simulation has 200 replications. The number of groups (N) takes values of 5, 30 and 500 while the number of periods (T) takes values of 5 and 20. This is to cover various ranges of panel width and length in empirical research.

Estimates for the coefficient β_1 of the time-varying variable x_1, and the coefficient γ_1 of the time-invariant variable z_1, using the FEVD estimator and the IV estimator are shown in Figures 3.1 and 3.2. As proved by Theorem 1, estimates by the

[4] The specified pattern of covariance is implemented through a Choleski decomposition approach.

two estimators, FEVD and IV, coincide *exactly* across all sampling distributions. Furthermore, increasing either width and length of a panel helps enhancing performance of both estimators by making their sampling distribution collapse around their expected values.[5]

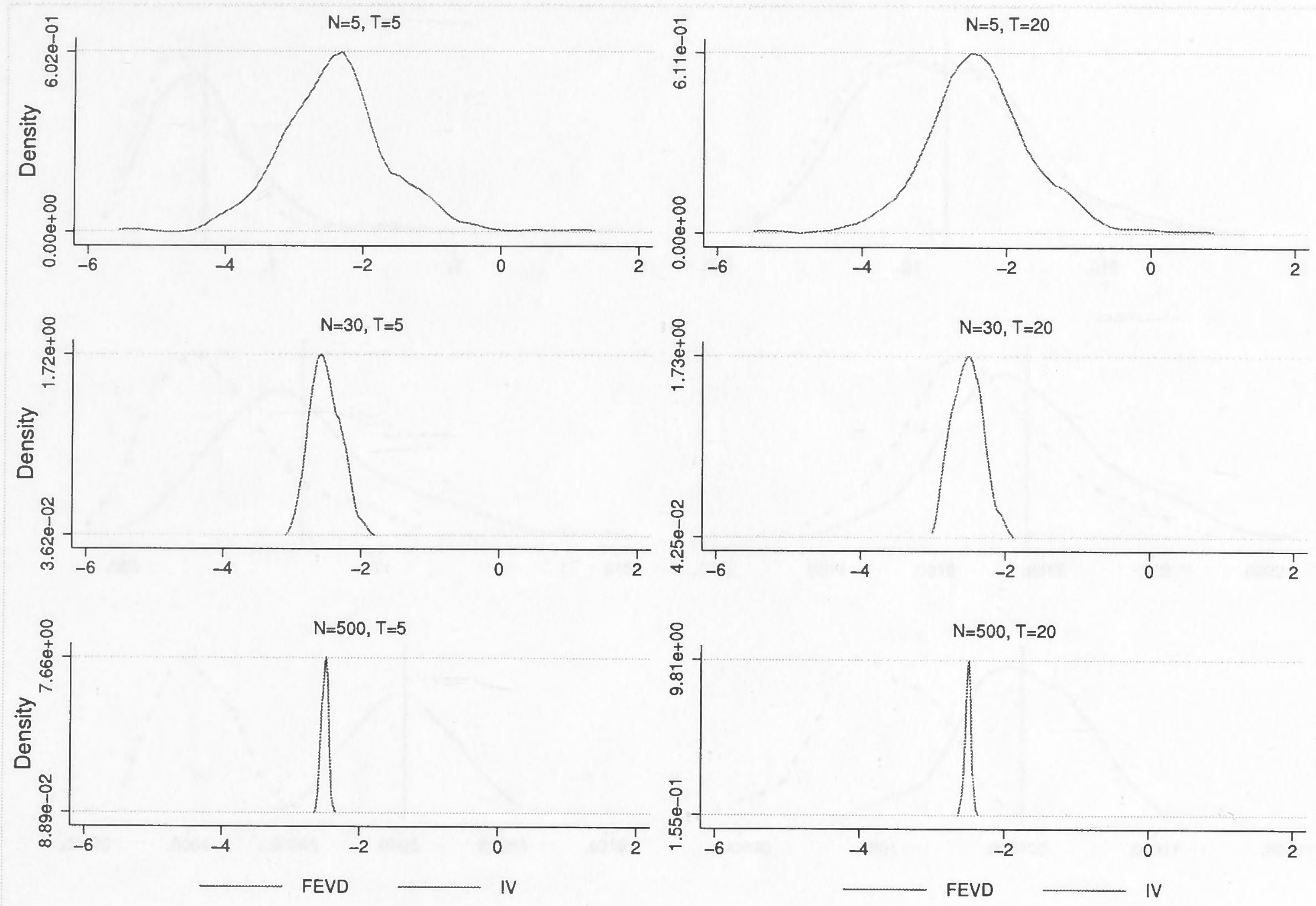
FIGURE 3.1: Sampling Distributions of beta 1



Sampling distributions of variances for β_1 and γ_1, using the FEVD estimator and the IV estimator, are shown in Figures 3.3 and 3.4, respectively. As shown in Figures 3.3 and 3.4, the means and distributions of estimates for variances of β_1 and γ_1, generated by using the IV variance formula, are always on the right of those generated by using the IV variance formula, regardless of the width and length of the panel data. Therefore, these figures collaborate our Theorems 2 and 3, which say that the FEVD variance formulas for both time-varying and time-invariant variables are *too small*.[6]

[5]Similar results for estimates for parameters β_2 and γ_2 are presented in the Appendix 3.6.3.
[6]Similar results for estimates of variances for β_2 and γ_2 are presented in the Appendix 3.6.3.

FIGURE 3.2: Sampling Distributions of gamma 1



Monte Carlo Simulation with 200 replications; Kernel = epanechnikov.

3.4 Comparison to Alternative Estimators

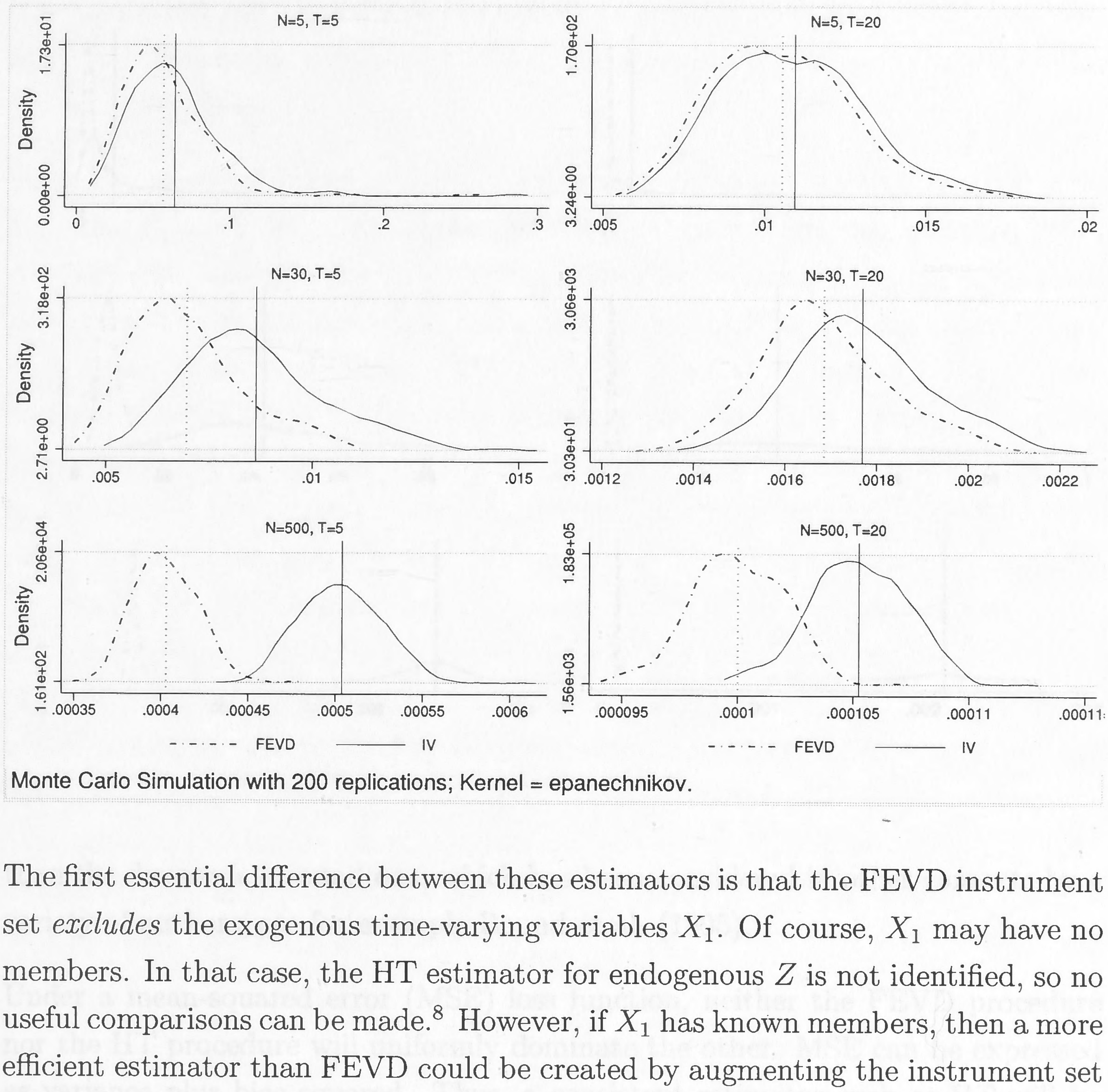
3.4.1 Theory Framework

The FEVD estimator was introduced as an alternative to the HT instrumental variable estimator. By also expressing FEVD in its instrumental variable representation we are able to develop insights into their comparative properties. Hausman and Taylor showed that the standard fixed effects estimator is equivalent to an IV estimator with instrument set $Q_D X$. To that, they add any exogenous elements of X or of Z as further instruments.[7]

To see the relationship more clearly, decompose X and Z into exogenous and potentially endogenous sets: $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$, where the subscript 1 indicates exogenous variables and the subscript 2 indicates endogenous variables. The HT procedure is then an IV estimator which uses the instrument set $[Q_D X, X_1, Z_1]$. In contrast, the FEVD procedure is an IV estimator which uses the instrument set $[Q_D X, Z_1, Z_2]$.

[7]Hausman and Taylor describe $P_D X$ as the additional instrument, but this interpretation follows Breusch et al. (1989).

FIGURE 3.3: Sampling Distribution of Variance for beta 1



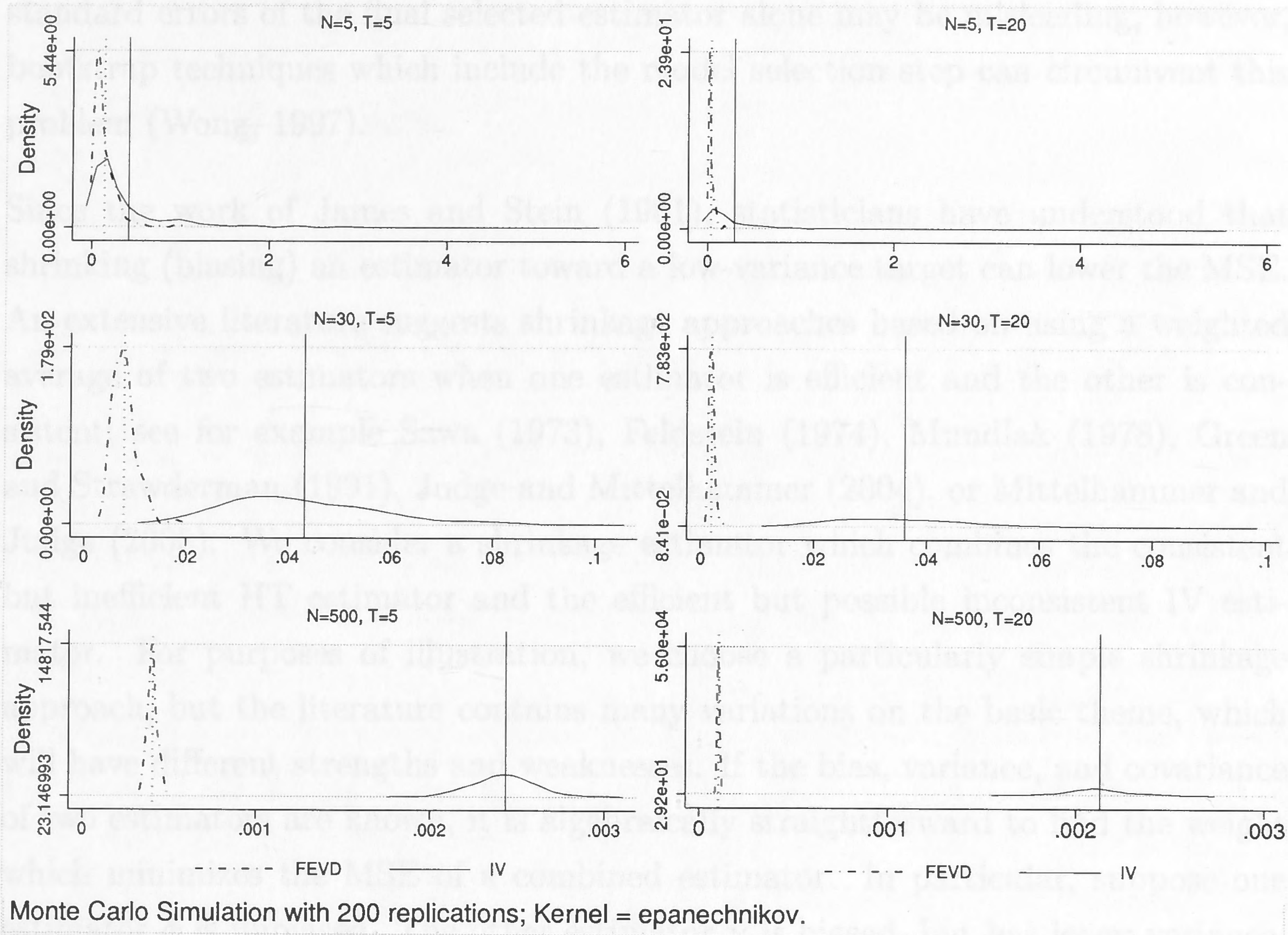
The first essential difference between these estimators is that the FEVD instrument set *excludes* the exogenous time-varying variables X_1. Of course, X_1 may have no members. In that case, the HT estimator for endogenous Z is not identified, so no useful comparisons can be made.[8] However, if X_1 has known members, then a more efficient estimator than FEVD could be created by augmenting the instrument set with X_1. The second essential difference is that the FEVD instrument set *includes* the potentially endogenous time-invariant variables Z_2. If these variables are in fact correlated with the group effect, then the FEVD estimator is inconsistent.

The FEVD and HT estimators coincide exactly when there are no exogenous elements of X and no endogenous elements of Z.[9] The FEVD procedure is thus primarily of interest when some Z may in fact be endogenous. The essential question raised by Plümper and Troeger is then whether it is better to use a biased and inconsistent but lower-variance estimator, or a consistent but higher-variance estimator. The question of whether a weak-instruments cure is worse

[8]Ideally, one would have theoretical grounds for identifying which elements of X are exogenous. As a practical matter, one could also use an over-identification test to confirm this assumption, since the fixed effects estimator of β is consistent.

[9]More precisely, the two estimators are identical when all elements of X are treated as if endogenous and all elements of Z are treated as if exogenous, regardless of the actual endogeneity status. See section 3.3.4 for an illustration.

FIGURE 3.4: Sampling Distribution of Variance for gamma 1



than the disease is a sound one, which has been considered in other contexts by a variety of authors; see for example Bound et al. (1995).

Under a mean-squared error (MSE) loss function, neither the FEVD procedure nor the HT procedure will uniformly dominate the other. MSE can be expressed as variance plus bias squared. Thus, a consistent estimator such as HT will be preferable to the FEVD for sufficiently large sample size.[10] In contrast, for a small sample with a small endogeneity problem, it might be preferable to include the the time-invariant endogenous variables Z_2 as instruments, as FEVD does. A more efficient estimator of this type than FEVD would be the IV estimator which augments the set of all valid instruments with Z_2, forming the instrument set $[Q_D X, X_1, Z]$.

One conventional approach to finding a balance would be to select between the competing estimators based on a specification test (Baltagi et al., 2003). If the test rejects the null hypothesis of no difference between estimators, then HT would be selected. Otherwise, the efficient estimator estimator would be selected because the evidence of endogeneity is too weak. Selection of a final estimator based on the

[10] Of course, consistency does require that valid instruments correlated with Z_2 exist.

results of a preliminary test is known as a *pretest* procedure. Inference based on the standard errors of the final selected estimator alone may be misleading; however, bootstrap techniques which include the model selection step can circumvent this problem (Wong, 1997).

Since the work of James and Stein (1961), statisticians have understood that shrinking (biasing) an estimator toward a low-variance target can lower the MSE. An extensive literature suggests shrinkage approaches based on using a weighted average of two estimators when one estimator is efficient and the other is consistent; see for example Sawa (1973), Feldstein (1974), Mundlak (1978), Green and Strawderman (1991), Judge and Mittelhammer (2004), or Mittelhammer and Judge (2005). We consider a shrinkage estimator which combines the consistent but inefficient HT estimator and the efficient but possible inconsistent IV estimator. For purposes of illustration, we choose a particularly simple shrinkage approach, but the literature contains many variations on the basic theme, which will have different strengths and weaknesses. If the bias, variance, and covariance of two estimators are known, it is algebraically straightforward to find the weight which minimizes the MSE of a combined estimator. In particular, suppose one estimator ϕ is unbiased. The other estimator χ is biased, but has lower variance. The shrinkage estimator then has the form $\chi + w(\phi - \chi)$, where w is the weight placed on the consistent estimator. Straightforward calculus shows that optimal weight which minimizes MSE is

$$w = \frac{\mu_\chi^2 + \sigma_\chi^2 - \sigma_{\chi\phi}}{\mu_\chi^2 + \sigma_\chi^2 + \sigma_\phi^2 - 2\sigma_{\chi\phi}}, \tag{3.24}$$

where bias is indicated by μ and where variance is indicated by σ.

Of course, the exact bias and variances will usually not be known; however, practical estimates of these terms are readily available for IV estimators. Mittelhammer and Judge (2005) show that plugging in such empirical estimates produces a practical weighted-average estimator. They choose a single w to minimize the sum of MSE over all coefficients. Since we are primarily interested in the MSE of a single coefficient in this analysis, we apply the solution for w, as presented in (3.24) which is the single-covariate case of equation 3.5 of Mittelhammer and Judge. We use standard empirical estimates of the variance and covariance terms from application of the basic IV formula (3.14). The difference between the two estimators provides our estimate of the bias of the efficient estimator, since HT is asymptotically unbiased. Mittelhammer and Judge provide detailed discussion on calculating bootstrap percentiles and standard errors, through application of

Efron's bias-corrected and accelerated bootstrap (Efron, 1987). The only change needed for the present context is to account for the panel structure, which is most simply done by resampling at the group level rather than resampling single observations independently.

3.4.2 Monte Carlo Evidence for Comparing the $FEVD$ to Alternative Estimators

In this section we compare the practical performance under a range of conditions of various estimators for an endogenous time-invariant Z. In addition to the FEVD and HT estimators, we consider a pretest estimator and a shrinkage estimator. The pretest estimator selects between HT and the IV estimator based on the instrument set set $[Q_D X, X_1, Z]$, which treats all Z as exogenous (as FEVD does) in addition to using the HT instruments. The pretest selection is based on the 95% critical value of the Durbin-Wu-Hausman specification test for exogeneity of Z (see *e.g.* Davidson and MacKinnon, 1993, p. 237). The shrinkage estimator assigns weights for the two estimators according to a first-stage empirical estimate of formula (3.24).

Plümper and Troeger argue for the superiority of the FEVD procedure over the HT approach based on Monte Carlo evidence. While our simulation design stays close to the original design where appropriate, our design differ from theirs in two fundamental respects.[11] The first difference is that in the Plümper and Troeger Monte Carlo study, the HT estimator was not actually consistent. This is because their data generating process had no correlation between X and Z. The fact that the available instruments had, by construction, zero explanatory power for the endogenous variable contrasts sharply with their characterization of the Monte Carlo results (p. 130): "the advantages of the FEVD estimator over the Hausman-Taylor cannot be explained by the poor quality of the instruments." Plümper and Troeger note (in footnote 11) that the advantage of FEVD persists in their experiments regardless of sample size. However, the asymptotic bias of an IV estimator is the same as the bias of OLS when the instruments are uncorrelated with the endogenous variable, and thus irrelevant (Han and Schmidt, 2001). In contrast, with a valid and relevant instrument, the bias of the IV estimator will approach zero asymptotically. We therefore consider scenarios in our simulation

[11] The authors graciously provided the original simulation code upon request.

where the HT estimator is consistent, that is at where at least one instrument for the endogenous Z is valid and relevant.

The second difference is that our simulations account for random variation in the group effect, while the Plümper and Troeger code holds the effect (u) fixed across all replications. Mundlak (1978) shows there is no loss of generality in assuming the effect is random, because the fixed-effects estimator and its related procedures can be described as inference conditional on the realizations of the effect in the sample. Further, the effect needs to be at least potentially random if the relationship between the effect and the regressors is to be described as *correlation.* As Mundlak shows, if the random effect is correlated with the group-averages of regressors in unknown ways, then the optimal linear estimator in the random-effects *model* is in fact the fixed-effects *estimator.*

The code used by Plümper and Troeger does not simply fix the replicated effects at some sample realization, rather it uses the Stata command 'corr2data' to fix the sample moments of the variables and the group effects exactly in every replication. The vector of effects is thereby 'fixed' by making it exactly orthogonal to the exogenous variables, effectively excluding any practical influence of the group effect in the simulated data. That process does not simulate a fixed-effects model, but rather one in which there is no group effect at all! By contrast, our random-effects simulation represents the situation where the analyst is uncertain of the magnitudes of the group effects.

We run a series of experiments which vary the degree of endogeneity and strength of instrument. The data generating process for our simulation is

$$y_{it} = 1 + 0.5x_1 + 2x_2 - 1.5x_3 - 2.5z_1 + 1.8z_2 + 3z_3 + u_i + e_{it}. \tag{3.25}$$

Here, $[x_1, x_2, x_3]$ is a time-varying mean-zero orthonormal design matrix, fixed across all experiments. $[z_1, z_2]$ is a time-invariant mean-zero orthonormal design matrix, fixed across all experiments. z_3 is fixed for all replications in each experiment. z_3 has sample mean zero and variance 1, and is orthogonal to all other variables except x_1. The sample covariance of the group mean of x_1 with z_3 is set exactly to an experiment-specific level, which allows us to vary the strength of the instrument across experiments.[12] The idiosyncratic error term e is standard normal. The random effect u is drawn from a normal distribution in each replication. The expectation of u conditional on z_3 is ρz_3, where ρ works out to be the value of

[12]Conditional on a non-zero sample correlation of the endogenous variable and the instrument, the moments of the IV estimator exist, so the Monte Carlo MSE is well-defined.

$\text{cov}(z_3, u)$ set in the experimental design. All other variables are uncorrelated with u, and the variance of u conditional on all variables is 1. The level of endogeneity is varied across experiments by changing the value of $\text{cov}(z_3, u)$. Each experiment has 1000 replications, which vary the random components u and e. There are 30 groups (N) and 20 periods (T), as reported in Plümper and Troeger (2007a). In implementing the estimators $[x_1, x_2, z_1, z_2]$ are treated as known exogenous, while $[x_3, z_3]$ are treated as potentially endogenous.

FIGURE 3.5: Performance of the Four Estimators for Varying Instrument Strengths

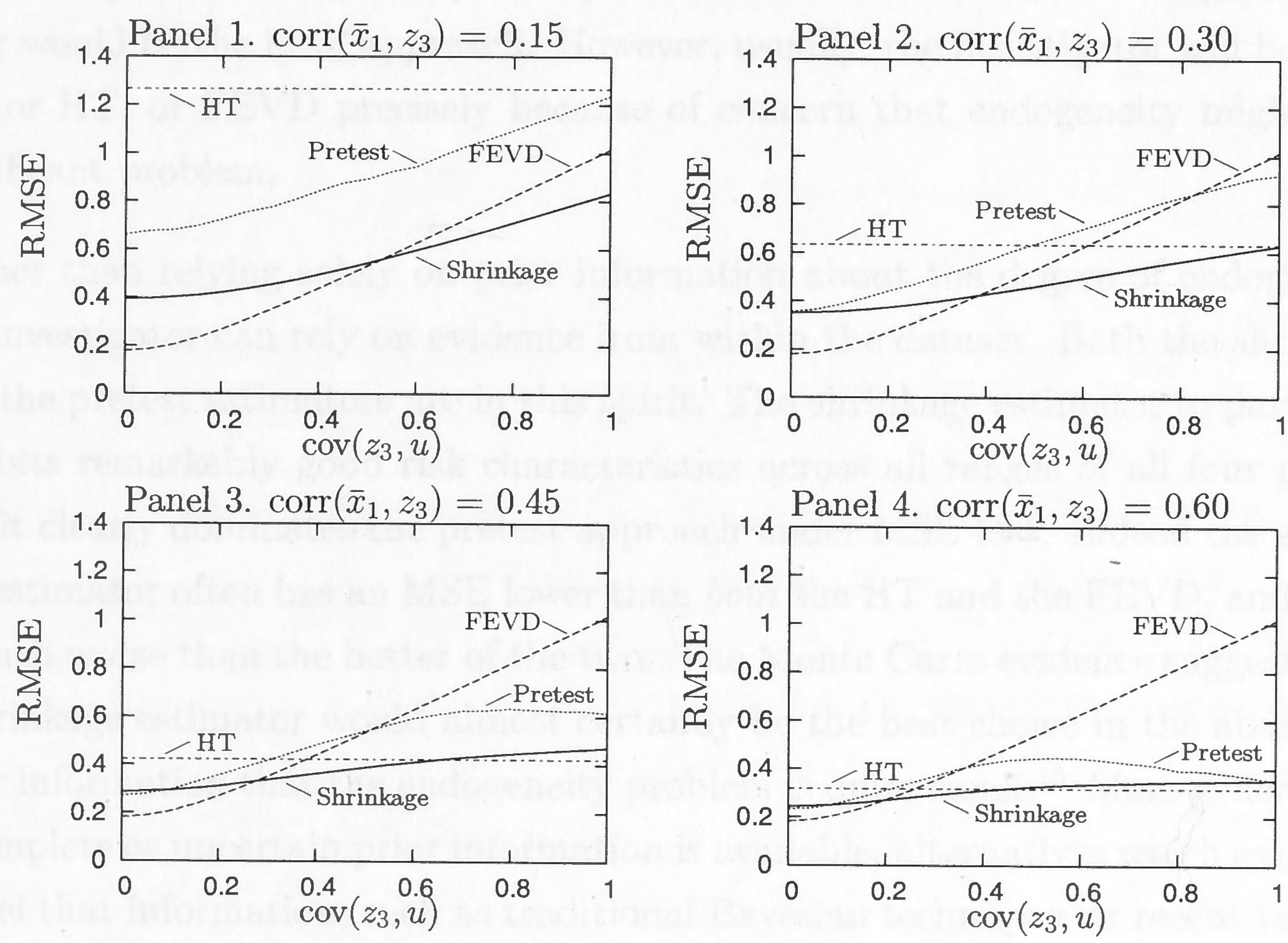


Figure 3.5 illustrates the simulation results for varying instrument strengths and endogeneity levels. The vertical axis in each panel is the square root of MSE of various estimators for the endogenous time-invariant variable z_3. The horizontal axis of each panel is the covariance between the random effect u and z_3. Each panel illustrates different instrument strength, as indicated by stronger instruments having higher correlation between the group-means of x_1 and the endogenous variable z_3. The four panels display the experiments for $\text{corr}(\bar{x}_1, z_3) = 0.15, 0.30, 0.45$, and 0.60 respectively.[13] Note that, within each panel, the HT results are unchanging as a consequence of the experimental design. Also, across panels, the FEVD results are unchanging by design.

[13]Because variances of $\bar{x}_1$ and z_3 are both 1, the covariance of these variables equals their correlation.

The most notable feature of Figure 3.5 is that neither HT nor FEVD uniformly dominates the other. If reasonably strong instruments are available to implement the HT procedure, and endogeneity is an issue, HT can greatly outperform FEVD as shown in Panel 4 because the higher variance of HT is compensated by lower bias.[14] For all cases when endogeneity is absent (or is mild), FEVD will be the most efficient estimator, as shown at the far left of all panels, because FEVD exploits the true (or approximately true) restriction that z_3 is uncorrelated with u. If the investigator has strong prior reason to believe that endogeneity is not an issue, it makes sense to use that information. Indeed, with informative priors over endogeneity, using a Bayesian procedure which minimizes risk against that prior would be the ideal approach. However, usually, the investigator will be using FE, or HT, or FEVD precisely because of concern that endogeneity might be a significant problem.

Rather than relying solely on prior information about the degree of endogeneity, the investigator can rely on evidence from within the dataset. Both the shrinkage and the pretest estimators are in this spirit. The shrinkage estimator in particular exhibits remarkably good risk characteristics across all ranges of all four panels, and it clearly dominates the pretest approach under MSE loss. Indeed the shrinkage estimator often has an MSE lower than *both* the HT and the FEVD, and never is much worse than the better of the two. The Monte Carlo evidence suggests that a shrinkage estimator would almost certainly be the best choice in the absence of prior information that the endogeneity problem is quite small.[15] More generally, if incomplete or uncertain prior information is available, alternatives which explicitly model that information, such as traditional Bayesian techniques or recent variants such as Bayesian model averaging (Hoeting et al., 1999), will likely be the best approach.

[14] The discussion here focuses on the small sample properties. When N is very large, HT will always outperform FEVD if there is endogeneity and valid and relevant instruments exist. For a modest example of relative estimator performance as N grows, see the Appendix 3.6.1, where the case of $N = 300$ and $T = 2$ is illustrated.

[15] While our focus is on estimator performance, it is worth noting that the Monte Carlo results do confirm that the asymptotic variance formula in (3.14) provides unbiased estimates of the HT and FEVD sampling variance, when σ_e^2 and σ_u^2 are calculated with appropriate degrees of freedom corrections for small sample. Further, the bootstrap quantiles for the shrinkage estimator are reasonably accurate, confirming the results of Mittelhammer and Judge (2005).

3.5 Conclusions

The FEVD estimator of Plümper and Troeger (2007a) offers the analyst of panel data a way to include time-invariant (and slowly-changing) variables in the presence of group effects that are possibly correlated with the explanatory variables. Thus it appears superior to the existing leading approaches of fixed-effects (which omits the time-invariant variables) and Hausman-Taylor (which requires specifying the exogeneity status of each explanatory variable). Plümper and Troeger's motivation for the procedure was mostly heuristic and their evidence came from Monte Carlo experiments showing that FEVD often displays better mean-squared error properties than both FE and HT. The procedure can be implemented in three easy stages, or even more conveniently in the Stata package provided by Plümper and Troeger (2007b). This procedure has proved popular with panel data analysts.

Our analytical results and revised Monte Carlo experiments challenge the value of FEVD. Is it still a useful tool?

We find that the coefficients of all the time-varying variables after the three stages of FEVD are exactly the same as FE in the first stage. This fact is sometimes seen in the empirical applications but rarely commented upon with any clarity. Obviously, there is no gain in using FEVD over the simpler FE if these coefficients are the objects of interest. Further, if something is known about the exogeneity of explanatory variables then these estimates are inefficient because they ignore the extra information. What is worse, unlike the simple first-stage FE, the standard errors from FEVD are too small — sometimes very much too small, judging from our empirical example and other published applications. In this case FEVD is a definite step backwards.

The main attraction of FEVD is its ability to estimate coefficients of time-invariant explanatory variables. But, again the third stage is questionable. The same coefficient estimates are given in the second stage, which is a simple regression of the group-averaged residuals from FE on the time-invariant variables. The purported value of the third stage is to correct the standard errors, but this reasoning is now known to be false. Indeed there will be cases where the second-stage standard errors — even though they are known to be wrong — will be more accurate than those from the third stage. The example we have provided in Section 3 shows this possibility.

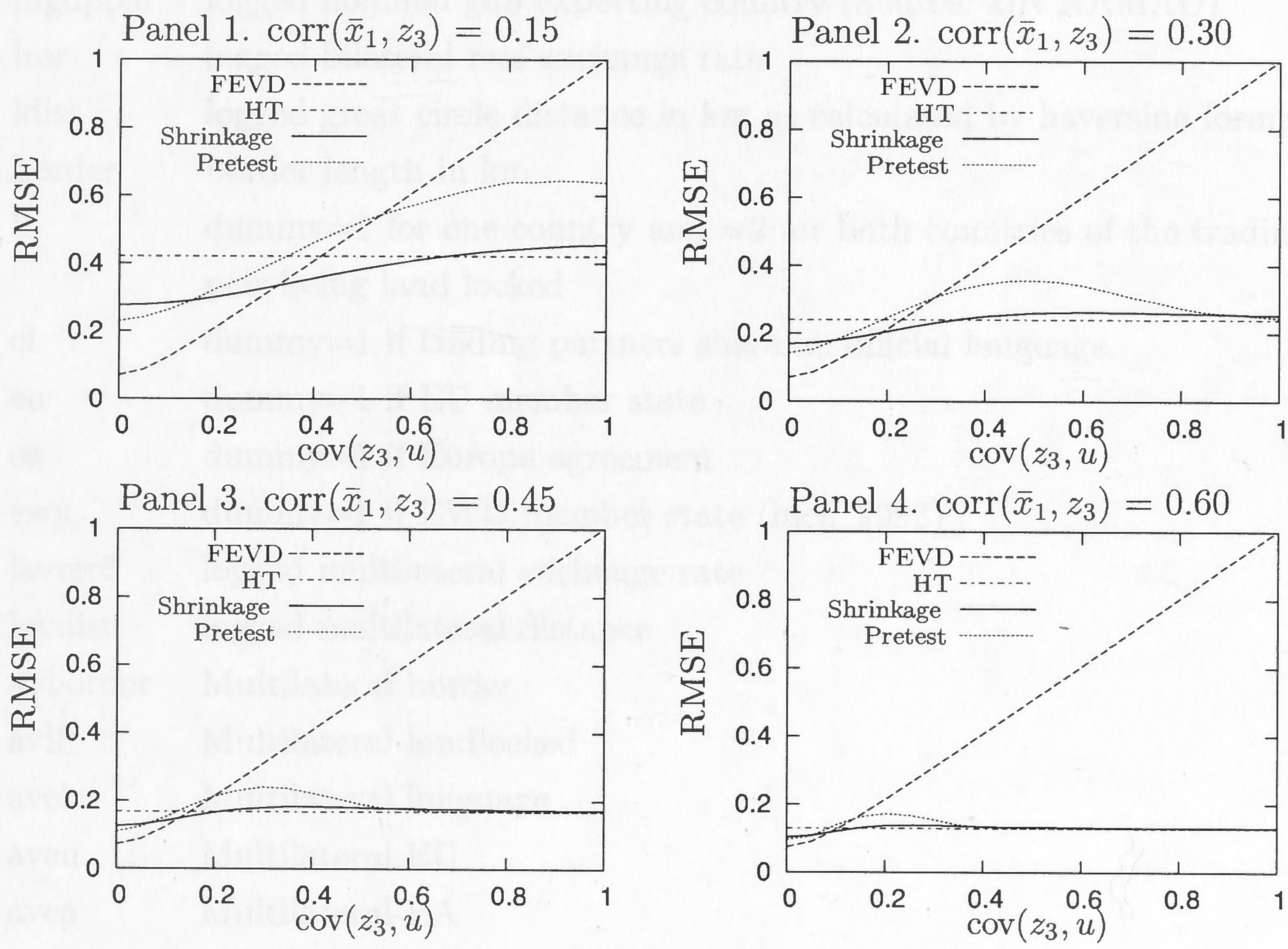
So if FEVD is the label to describe the three-stage procedure, it cannot be recommended for making inferences about any of the coefficients. The coefficient estimator, however, also represents a particular choice of instruments in standard IV . Dropping the three-stage methodology and reverting to an explicit IV approach would allow correct standard errors to be obtained in the cases where the estimator is consistent. However, since all of the time-invariant variables are used as instruments, the FEVD estimator will be inconsistent if any of these are endogenous. The value of this estimator relative to others then depends on the trade-off between inconsistency and inefficiency.

When the objective is reduced mean-squared error, the literature is replete with other methods such as shrinkage estimators known to have good properties. We have provided one such estimator that clearly dominates the FEVD estimator over much of the parameter space and also limits the risk in regions where the FEVD risk is unbounded. We demonstrate the feasibility of such an estimator with standard errors found empirically by bootstrapping. In undertaking these investigations we have also uncovered an explanation for the misleading evidence favouring FEVD that was suggested in the previous Monte Carlo studies.

3.6 Appendices

3.6.1 Monte Carlo Evidence for Comparing $FEVD$ to Alternative Estimators for Large N and Small T.

FIGURE 3.6: Relative Estimator Performance when $N = 300$ and $T = 2$.



In applications such as labor market studies the number of groups can be quite large, often in the tens of thousands, since there may be a distinct group for each individual in the study. Figure 3.6 presents a modest example of the relative behavior of the four estimators as the number of groups grows larger. Each panel in Figure 3.6 illustrates the same parameter settings as the corresponding panel in Figure 3.5. The simulation code for the figures is identical, except for the N and T settings. While the overall number of observations is the same in the two figures, the larger number of groups provides more information about the time-invariant variables. Panel 4 illustrates that the relative performance of FEVD can be quite poor for reasonable parameter settings and a modest number of observations.

3.6.2 Full Results for the Empirical Evidence for the Theorems

Variable Descriptions

lnimports	logged imports in 1000 US$ (Source: OECD ICTS)
lngdprep	logged nominal gdp importing country (Source: UN NAMAD)
lngdppar	logged nominal gdp exporting country (Source: UN NAMAD)
lrer	logged bilateral real exchange rate
ldist	logged great circle distance in km as calculated by haversine formula
border	border length in km
ll	dummy=1 for one country and =2 for both countries of the trading pair being land locked
cl	dummy=1 if trading partners share an official language
eu	dummy=1 if EU member state
ea	dummy=1 if Europe agreement
ewu	dummy=1 if EWU member state (incl. 1998)
lavrer3	logged multilateral exchange rate
lavdist	logged multilateral distance
avborder	Multilateral border
avll	Multilateral landlocked
avcl	Multilateral language
aveu	Multilateral EU
avea	Multilateral EA
avewu	Multilateral EWU
hc1	number of years of a trading pair in the sample
hc3	dummy=1 if the trading pair is present in the sample in t-1

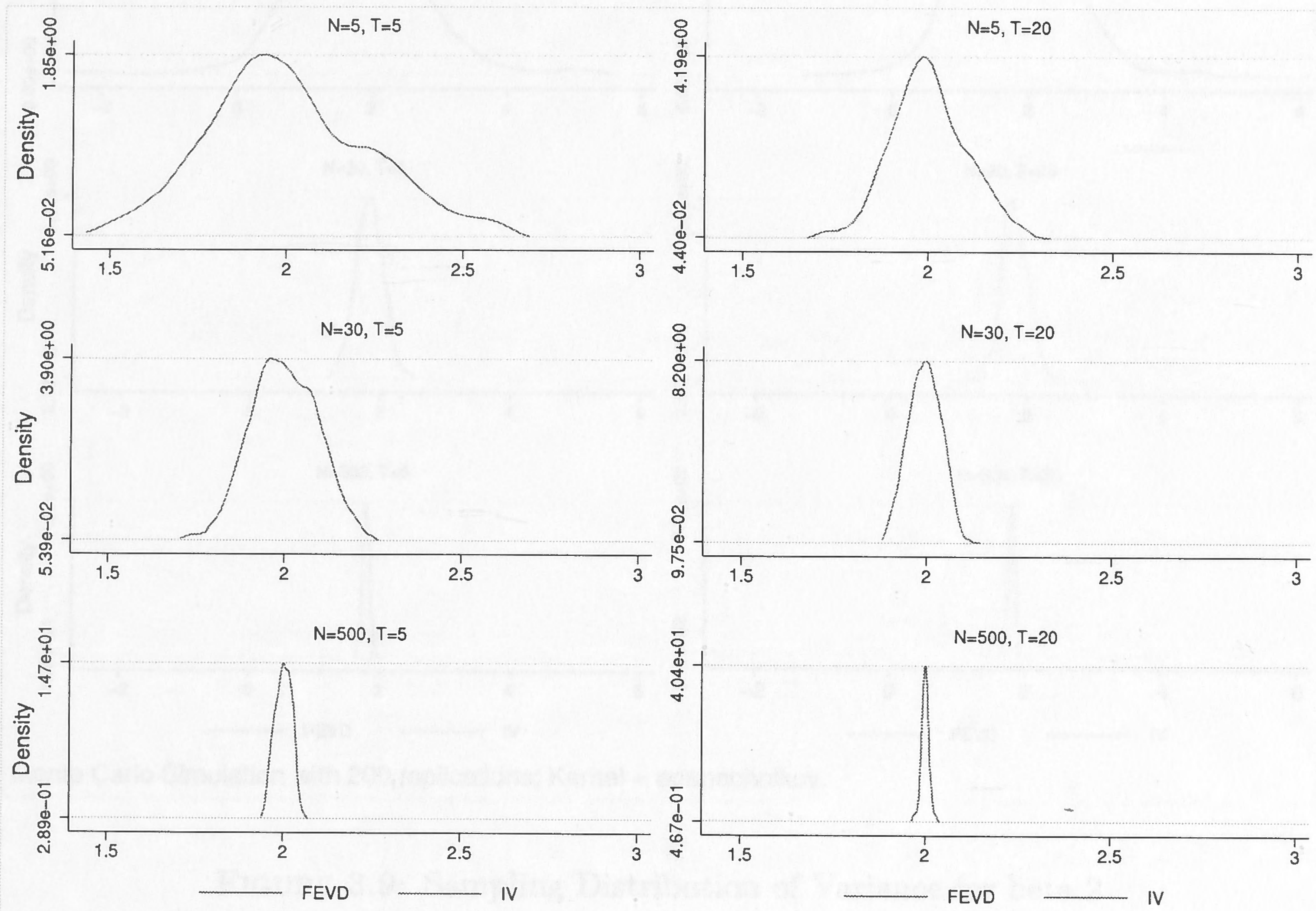
Table 3.2: Full Replication of Belke and Spies (2008).

	(1) POLS	(2) FE	(3) Stage 2	(4) FEVD	(5) HT
lngdpim	0.88***	0.68***	.	0.68***	0.68***
	(0.04)	(0.11)	.	(0.01)	(0.11)
lngdpex	0.89***	0.71***	.	0.71***	0.71***
	(0.03)	(0.07)	.	(0.01)	(0.07)
lrer	-0.01	0.13**	.	0.13***	0.13**
	(0.01)	(0.06)	.	(0.01)	(0.06)
ldist	-1.27***	.	-1.41***	-1.41***	-1.75***
	(0.11)	.	(0.04)	(0.00)	(0.16)
border	-0.00	.	0.00**	0.00***	-0.00
	(0.00)	.	(0.00)	(0.00)	(0.00)
ll	-0.16*	.	-0.23***	-0.23***	-0.16
	(0.10)	.	(0.04)	(0.00)	(0.14)
cl	0.23*	.	0.13***	0.13***	0.02
	(0.12)	.	(0.05)	(0.00)	(0.16)
eu	0.08	0.03	.	0.03***	0.03
	(0.09)	(0.05)	.	(0.01)	(0.05)
ea	0.16*	0.22***	.	0.22***	0.22***
	(0.10)	(0.06)	.	(0.01)	(0.06)
ewu	0.13**	0.07**	.	0.07**	0.07**
	(0.05)	(0.03)	.	(0.00)	(0.03)
lavrer3	1.22***	0.45**	.	0.45***	0.45***
	(0.41)	(0.22)	.	(0.03)	(0.22)
lavdist	0.55***	.	0.93***	0.93***	1.45***
	(0.15)	.	(0.04)	(0.00)	0.24)
avborder	0.00***	.	0.01***	0.01***	0.01***
	(0.00)	.	(0.00)	(0.00)	(0.00)
avll	-0.10***	.	-0.14***	-0.14***	-0.18***
	(0.03)	.	(0.01)	(0.00)	(0.06)
avcl	-0.02	.	-0.40***	-0.40***	-0.44
	(0.26)	.	(0.10)	(0.00)	(0.39)
aveu	-0.74***	-0.22*	.	-0.22***	-0.22*
	(0.21)	(0.12)	.	(0.01)	(0.12)
avea	0.34	-0.07	.	-0.07***	-0.07
	(0.23)	(0.10)	.	(0.01)	(0.10)
avewu	0.22*	0.69***	.	0.69***	0.69***
	(0.12)	(0.08)	.	(0.01)	(0.08)
hc1	0.09***	.	0.10***	0.10***	0.09*
	(0.03)	.	(0.01)	(0.00)	(0.05)
hc3	-0.18***	-0.03	.	-0.03***	-0.03
	(0.04)	(0.03)	.	(0.00)	(0.03)
δ	.	.	.	1.00***	.
	.	.	.	(0.00)	.

One, two, and three asterisks reflect significance at the 0.10, 0.05, and 0.01 confidence levels, respectively. Robust standard errors are in parentheses.

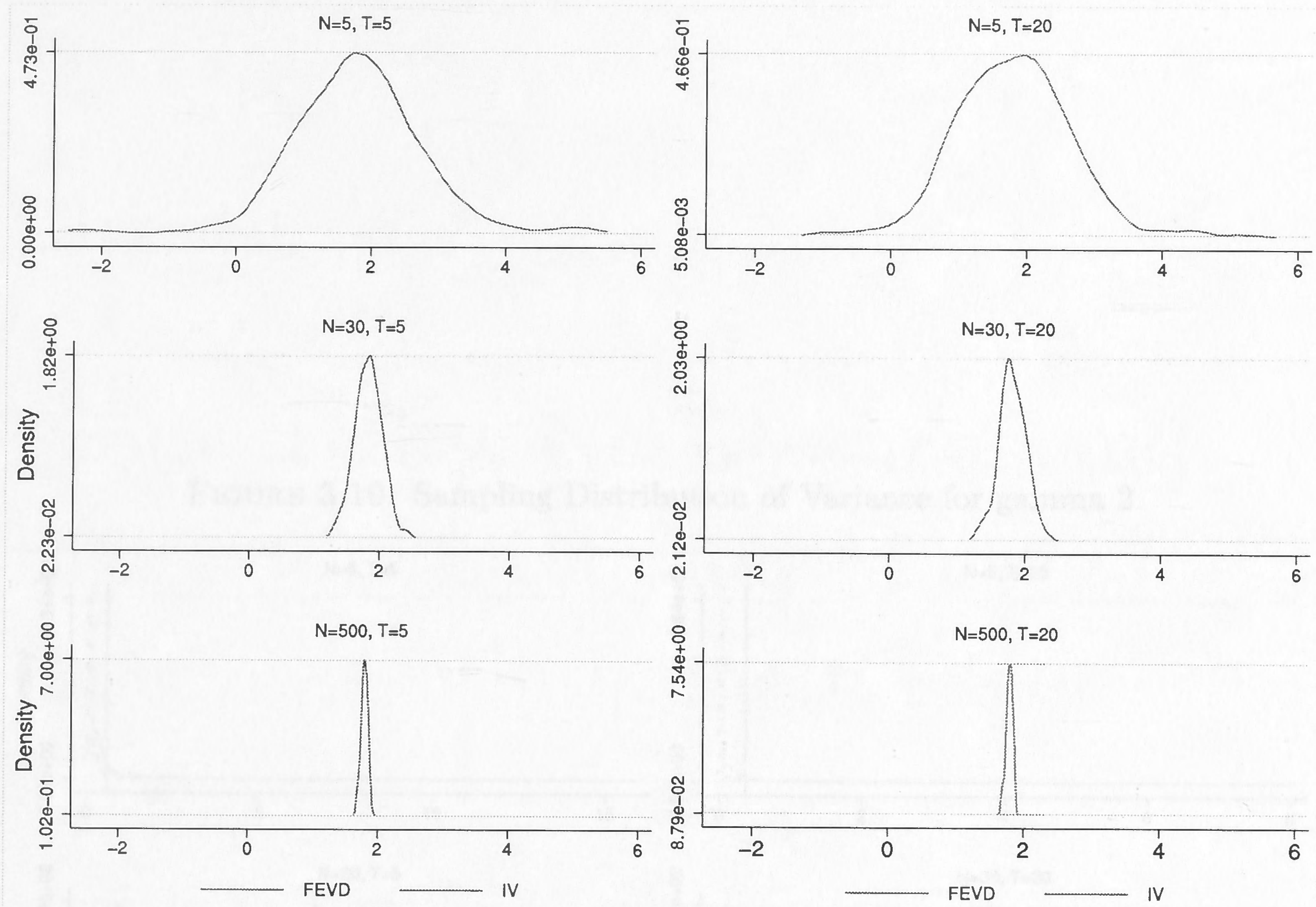
3.6.3 Full Results for the Monte Carlo Evidence for the Theorems

FIGURE 3.7: Sampling Distributions of beta 2



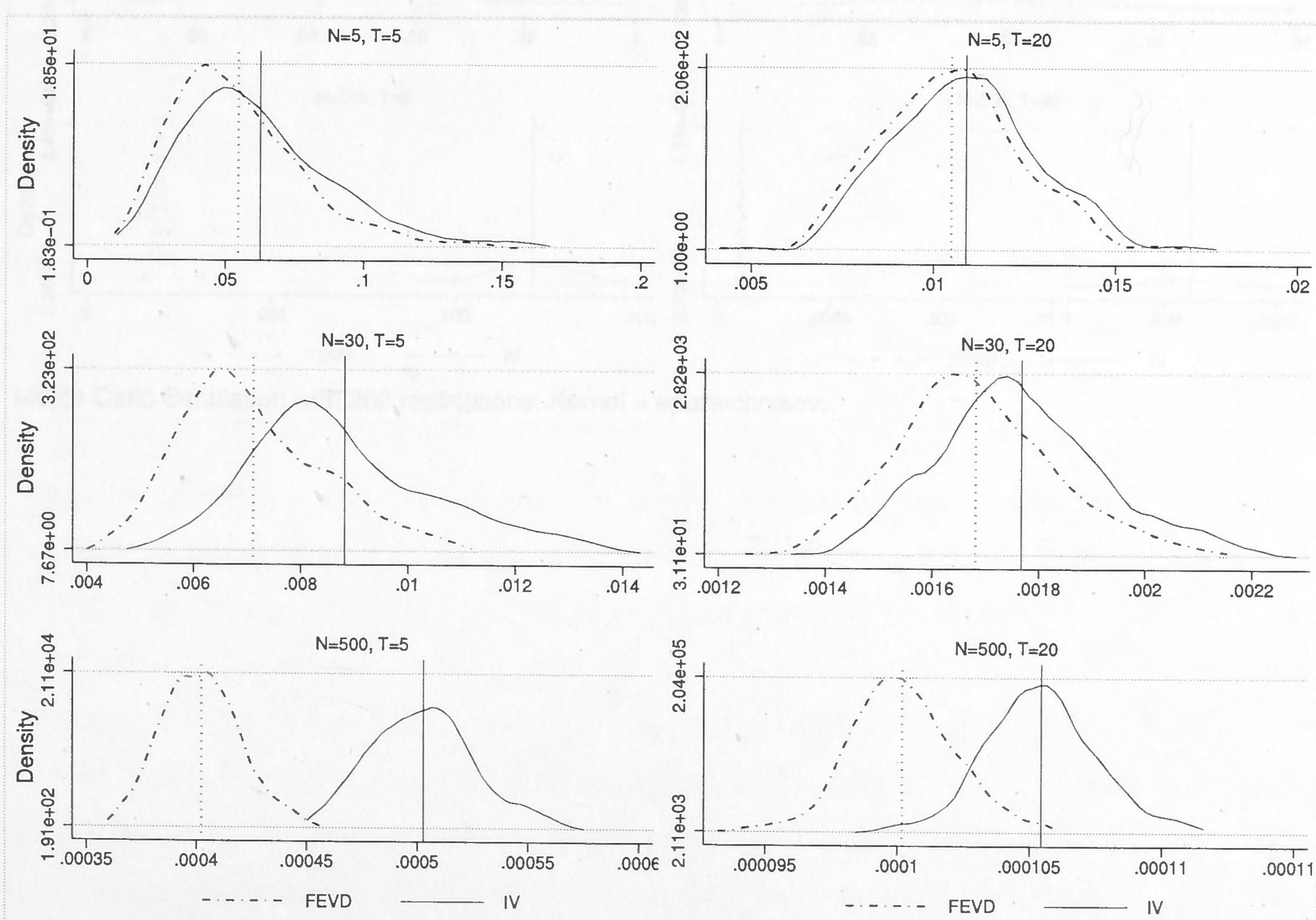
Monte Carlo Simulation with 200 replications; Kernel = epanechnikov.

FIGURE 3.8: Sampling Distributions of gamma 2



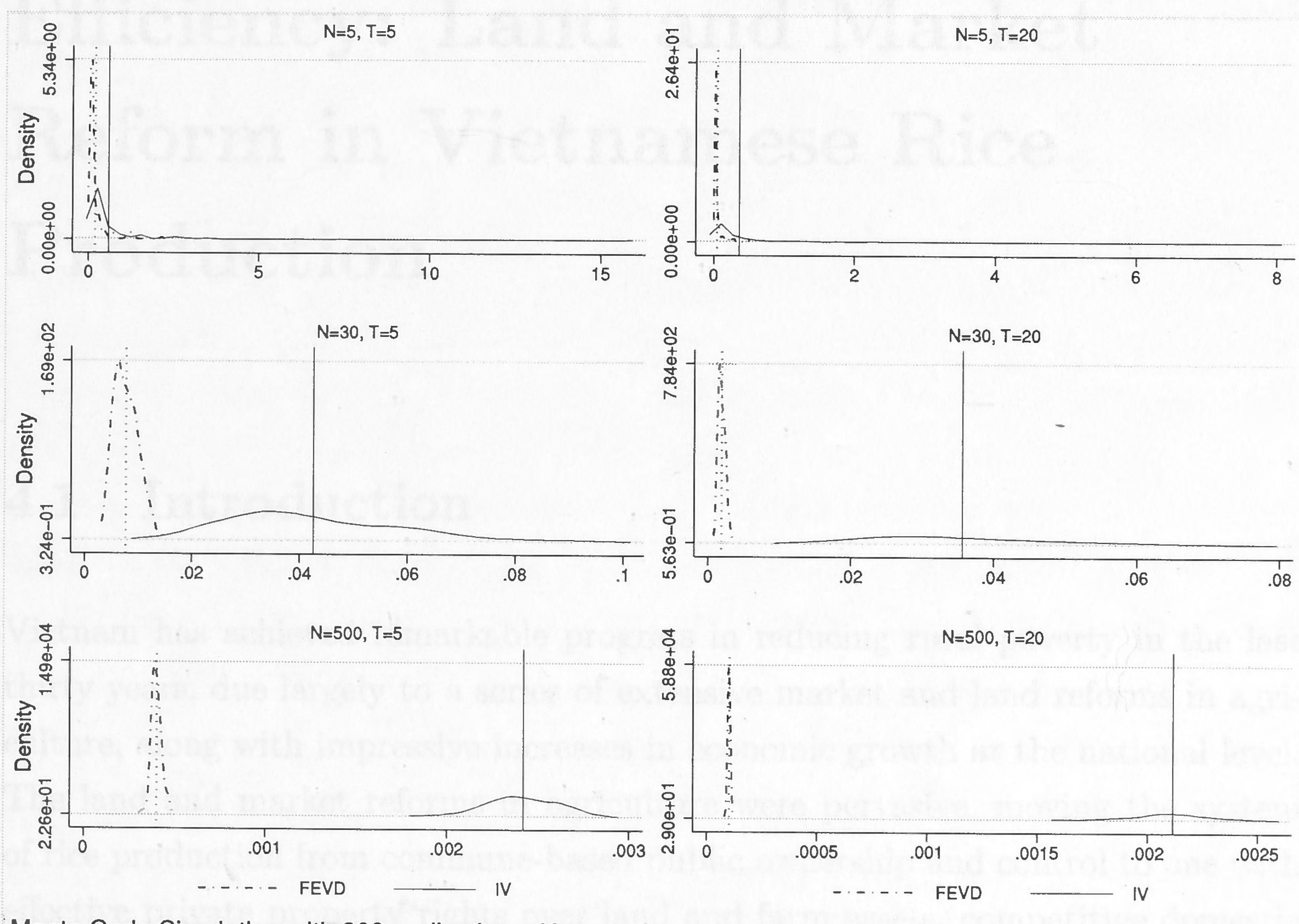
Monte Carlo Simulation with 200 replications; Kernel = epanechnikov.

FIGURE 3.9: Sampling Distribution of Variance for beta 2



Monte Carlo Simulation with 200 replications; Kernel = epanechnikov.

FIGURE 3.10: Sampling Distribution of Variance for gamma 2



Monte Carlo Simulation with 200 replications; Kernel = epanechnikov.

Chapter 4

Productivity, Net Returns and Efficiency: Land and Market Reform in Vietnamese Rice Production

4.1 Introduction

Vietnam has achieved remarkable progress in reducing rural poverty in the last thirty years, due largely to a series of extensive market and land reforms in agriculture, along with impressive increases in economic growth at the national level. The land and market reforms in agriculture were pervasive, moving the system of rice production from commune-based public ownership and control to one with effective private property rights over land and farm assets, competitive domestic markets and individual decision making over a wide range of agricultural activities. The substantial incentive effects created by these policy measures, inducing farmers to work harder and use land more efficiently, have been estimated to be as much as fifty percent of the increase in total factor productivity (TFP) during the peak of the reform period (Che et al., 2006). Overall, given these reforms, Vietnam has gone from being a large importer of rice from 1976–80, to now the second largest exporter of rice in the world, with considerable increases in farm profitability and rural incomes, and decreases in rural poverty rates by over forty percent from 1993 to 2004 alone (Hansen and Nguyen, 2007).

However, much still remains to be done to increase living standards in rural areas and enhance general rural development. Like many reform processes, the early rapid gains in economic activity have dissipated over time, with the suggestion now of a TFP 'slow down' in rice production in many areas in Vietnam. In addition, many of the poor still farm small areas of land, constrained in use, often with fragmented or non-contiguous plots, and with little or no human and physical capital accumulation or access to agricultural extension services and farm credit. Much of this is due to remnants from past institutional arrangements, but also to continued constraints in land use, credit availability and the provision of rural services, all calling for further or renewed land policy and market reform.

This paper has two basic tasks. First, it assembles a data set from 1985 to 2006 to measure the changes in TFP, terms of trade, net incomes and net returns in Vietnamese rice production, both in the principal rice growing areas and throughout the country. The results track the effects of the major land and market reform process and determine key differences in TFP and net returns over time and by region. All of this speaks directly to existing land use practices and is suggestive of needed policy response. The second task is to isolate the remaining institutional constraints and policy challenges that may be limiting increases in productivity and efficiency. For this purpose, two stochastic production frontier and inefficiency models, drawn from two different farm survey data sets, are estimated to determine the potential effects of ongoing issues over land use and sale, the provision of credit, land fragmentation, less than secure property rights and the lack of rural education and support services.

Section 2 of the paper provides context, highlighting the nature and extent of the past market and land reform process and the remaining institutional barriers and policy challenges in land use practice and rural development. Section 3 briefly summarizes the various data sets used in the paper, along with variable definitions and econometric specification. An extensive data appendix details the sources of the data, as well as data constructions and various adjustments for TFP calculations. Section 4 provides the measures of TFP, terms of trade, net incomes and net returns in rice production, while section 5 provides estimates for the stochastic frontier and inefficiency models. Section 6 concludes.

4.2 Context

Rice continues to dominate agricultural production in Vietnam, with rice production accounting for nearly 90 percent of the output of food grains and almost two-thirds of rural farm income (SDAFF, 2006). Although rice is produced in every one of the 60 (recently redefined as 64) provinces in Vietnam, the Red River Delta (RRD) and the Mekong River Delta (MRD) are the main rice growing regions. The smallest producers of rice (less than 100,000 tons per year) are in Binh Phuoc province, which is relatively small in area, and contains the principal coffee growing (Gialai Kontum) and mining (Cao Bang, Bac Kan) areas. Provinces with the largest rice output (more than a million tons per year) are located in the MRD (Tien Giang, Soc Trang, Long An, Kien Giang and An Giang), which as a whole accounts for roughly half of Vietnam's output of rice and most of its international exports. In terms of natural conditions, the MRD and the RRD are the most favorable for growing rice, with many areas naturally irrigated, producing up to three rice crops per year. Based on farm survey data for 2004 (as used below), the average farm size in the RRD (0.4 hectares per farm) is much smaller than in the MRD (1.4 hectares per farm). However, the number of threshing machines in the RRD is almost double that of the MRD. In the MRD, with a large volume of high-quality rice exports, rice processing takes place in mills rather than on the farm to maintain international standards.

Rice output has increased dramatically during the major land and market reform periods, from 10 million to nearly 34 million tons nationwide from 1980 to 2004 (Kompas, 2004). After a period of 'output share contracts' (from 1981–87), where up to 80 percent of rice output had to be delivered to the state, with the remainder sold in competitive markets, a period of 'trade liberalization' (1988–94) brought major institutional and market change, allowing for effective private property rights over both land (initially 10–15 and later 20 year leases) and capital equipment. With reform, most production decisions were decentralised, all farm income (after tax) was retained by the farmer and rice could be sold freely in competitive domestic markets. The result was an increase in rice prices (from state controlled low prices prior to reform), added profitability, considerable increases in TFP and dynamic gains from trade reform due to induced capital accumulation out of retained farm earnings (Che et al., 2001, 2006).

Since 1994, these dramatic market and land reforms have been solidified and in many cases extended in the 'post-reform' period.[1] Nevertheless, a number of concerns remain, and have been raised again recently in a 'Participatory Poverty Assessment' (PPA), with over 240 focus groups and 1,450 participants, undertaken by the Vietnam Academy of Social Sciences (Vietnam Academy of Social Sciences, 2009). Chief among these are issues surrounding land titles and their use, land fragmentation and the lack of rural credit availability and supporting rural services.

Land title and use requirements are a good example of the challenges that remain. Although the Land Law of 1993 (amended and clarified in 1998, 2001 and 2003) allows "land use rights to be transferred, exchanged, leased, mortgaged and inherited" (Congress of Vietnam, 1993), in practice, considerable constraints remain in place regarding both land conversion (i.e., land transferred or converted to other uses) and land accumulation (i.e., trades and accumulation of land plots). For any land conversion, for example, the commune authorities have to first develop a plan, often based on or as minor amendments to past historical 'blueprints' for land use in that area, to submit to various levels of government for approval. The PPA reports that this process is often long and transactions costs are high, making it difficult for poor farmers in particular to participate (Vietnam Academy of Social Sciences, 2009). In addition, although land consolidation in Vietnam is occurring, with a number of important benefits, there are still restrictions on overall land size (see Ravallion and Van de Walle, 2008). In 2007, the Vietnamese government increased limits on the transfer of land use rights for annual agricultural land from 3 to 6 hectares for the South East, the MRD and Ho Chi Minh City and from 2 to 4 hectares for other cities and provinces. This is a welcome albeit modest change for many farmers, but in most cases rice farming outside the MRD still takes place on very small farms, at subsistence levels (General Statistic Office, 2004; Vietnam Academy of Social Sciences, 2009).

Part of the obstacle to land consolidation is the lack of fully secure property rights. The terms of land use titles were extended for agricultural land from 15 to 20 years with the Land Law of 1993, but in many cases even 20 years is too short to provide secure rights in the shift to larger farms, or allow farm land to be turned into use for small manufacturing or industry. Overall the process of land certification or entitlement itself has also been below expectations. Household survey data from (General Statistic Office, 2006b) suggests that only 76 percent of agricultural land parcels, 68 percent of urban land parcels and 34 percent of forest land parcels have been granted land-use right certificates (World Bank, 2009b). Even where

[1] For a detailed discussion of land reform policy in Vietnam, see Chu et al. (1992); Fforde (1996); Marsh and MacAulay (2002)

land certificates do exist, there is a shortage of basic infrastructure for an effective operation of land administration, including cadastral mapping, transaction registrations and record management in the provision of land administration services. Problems are compounded by lack of information or public awareness and the apparent limited capacity of land administration staff, especially at the commune and district levels (World Bank, 2009b). It is perhaps for this reason that real estate markets have been slow to develop, with recent data indicating that the role of land rental markets in agriculture in rural areas remains thin (General Statistic Office, 2004), and that continued government restrictions often prevent low cost and efficiency enhancing transfers (Deininger and Jin, 2008). Land rights that are not secure also have an impact on credit availability and capital accumulation. The PPA reports that farms without land-use certificates, which would normally be used as collateral, are not able to obtain even short term loans, much less transfer land use entitlements. This is especially true for many farms with land tenures that are due to expire soon, given the 20 year leases initiated in 1993 (Vietnam Academy of Social Sciences, 2009).

Land fragmentation occurs when households have land use rights to a number of often small, non-contiguous plots. With reform and the resulting dissolution of the commune system in Vietnam, land was allocated to prior commune members in a roughly egalitarian manner: a more or less equal distribution of plots to each household, throughout the commune, so that no one household would have a concentration of plots in the most fertile parts of the commune, or near roads, water sources or other essential services. Immediately after the major reform process (1988–94), there were as many as 75 to 100 million parcels or plots of land in Vietnam and on average about seven to eight plots per household (World Bank, 2003a; Van Hung et al., 2007), of which about 10 percent had an area of only 100 square meters or less (Phien, 2001). Evidence suggests that total plot numbers have been falling recently with land consolidation (nationwide, falling to 4.3 plots per household (General Statistic Office, 2004), but the problem is still common. Fragmentation levels, for example, continue to remain high in the RRD, the most populated region, with 90 percent of the households having rice farm sizes of roughly 0.2 to 0.5 of a hectare (Chu, 2008) and an average of 4.6 plots per farm (General Statistic Office, 2004). The number of plots per farm in the MRD, by contrast, is only 1.6. In some cases (e.g., risk spreading), fragmentation may be an advantage (see Marsh et al., 2006). However, for the most part, small and scattered land holdings hamper mechanization, the use of buffaloes and tractors and technological adoption, and require additional time and labour for farming activities that must be carried out in geographically distant plots. The embankments that separate plots alone have been estimated to reduce total agricultural land for cultivation in Vietnam by 2.4 to 4 percent (Thanh, 2008).

4.3 Data Sources, Variables and Specifications

Three different data sets are used in this paper. The first is a provincial level data set on rice production in Vietnam, from 1985 to 2006, for 60 provinces, used to construct TFP, price and quantity indices and net income and return measures, greatly improving on the basic TFP estimates (to 1994) provided in (Che et al., 2006). This is an extensive data set on prices and quantities for all rice outputs and inputs, directly obtained as (or aggregated to) provincial averages. The key variables include paddy rice output, labour, land, material inputs (especially fertilizer, but also seeds and pesticide), capital (a measure of tractors and buffaloes), as well as input prices for labour, land, fertilizer, pesticides and capital used in rice production. (See the Appendix for detailed data sources, constructions and adjustments.)

The second data set, used to construct a stochastic production frontier and inefficiency model, is a farm survey data set for 388 rice farms, complied in 2004 (by the authors) in the major rice growing regions (the MRD and RRD), designed specifically to isolate the potential effects on inefficiency from differing farm characteristics and potential land fragmentation. Key additional variables include measures of soil quality and irrigation, average plot size, as a proxy for land fragmentation, and the level of education of the household head of the farm.

The third data set is the 2004 Vietnam Household Living Standards Survey (VHLSS), which is used to confirm and extend the results of the smaller farm survey data set, providing added estimates of the effects of secure property rights, more precise measures of the effects of land fragmentation (using a Simpson index) and access to agricultural extension services and credit. The VHLSS is a household survey data set of roughly 9,000 households in 2,216 communes, with cluster-sampling techniques to cover the entire country, conducted by the General Statistical Office (GSO) in Vietnam in selected years (e.g., 2002, 2004 and 2006). The 2004 VHLSS data set, in particular, compared to other VHLSS data sets in Vietnam, has an extended module for land, with separate components for land use and agricultural production. Sample size is reduced to 3,671 households to isolate farms that are primarily rice producers. It is important to note that a Simpson index is not applicable to the 2004 farm survey data, since the exact size of each plot for the 388 farm households was not recorded. The measure of fragmentation is thus simply total rice land divided by the number of plots.

For all stochastic production frontiers log-likelihood specification tests were used to determine functional form and the presence of inefficiency effects. In all cases, standard OLS estimates are seen to be inappropriate and functional form tests reject a translog specification in favor of a more standard Cobb-Douglas production function. Tests also indicate that estimates of the stochastic frontiers using a random coefficients approach, allowing for 'non-neutral' shifts in the production frontier, following Kalirajan and Obwona (1994), resulted in little difference in estimated coefficients, with the inefficiency term adequately represented by a truncated half-normal distribution. Frontier and inefficiency estimates are obtained using Frontier 4.1 (see Coelli et al., 1998).

4.4 Total Factor Productivity, Terms of Trade and Net Returns

The change in TFP is a measure of outputs to inputs over any two time periods. Results for Vietnamese rice production are generated using Tornqvist quantity (and price) indices given by

$$\ln Q_{st} = \sum_{i=1}^{N} \left(\frac{\omega_{is} + \omega_{it}}{2} \right) (\ln q_{it} - \ln q_{is}) \tag{4.1}$$

or

$$Q_{st} = \prod_{i=1}^{N} \left(\frac{q_{it}}{q_{is}} \right)^{\frac{\omega_{is}+\omega_{it}}{2}} \tag{4.2}$$

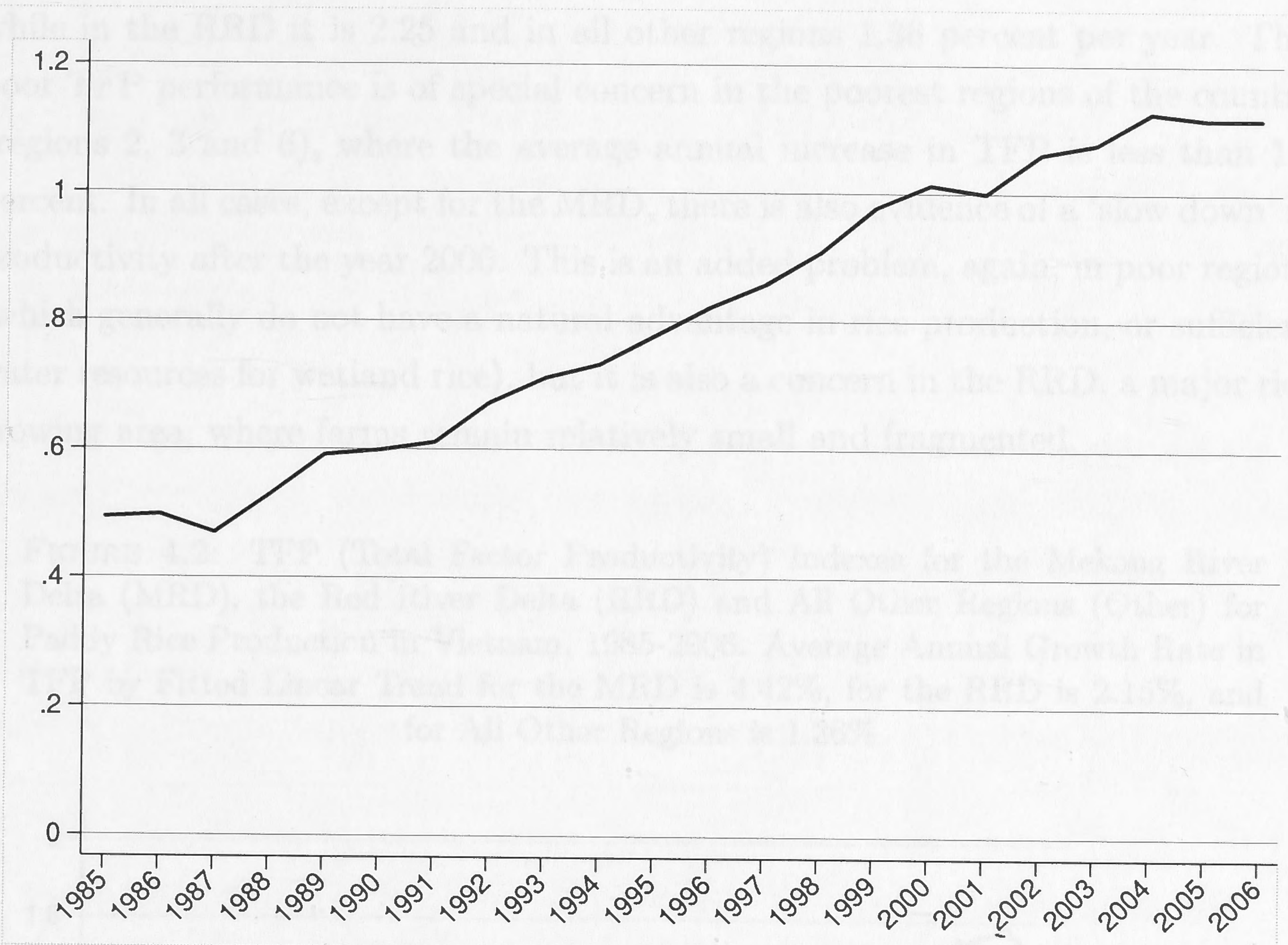
for N quantities, q inputs or outputs (depending on context), periods s and t and weights

$$\omega_{is} = p_{is} q_{is} / \sum_{i=1}^{N} p_{is} q_{is} \tag{4.3}$$

for time period s, for example. TFP, for outputs y and inputs x is thus given by

$$\ln TFP_{is} = \frac{1}{2} (\omega_{is} + \omega_{it}) (\ln y_{it} - \ln y_{is}) - \frac{1}{2} (v_{is} + v_{it}) (\ln x_{it} - \ln x_{is}) \tag{4.4}$$

FIGURE 4.1: Paddy Rice Output (Indexed) in Vietnam (1985-2006), Average Annual Growth Rate by Fitted Linear Trend = 3.5%

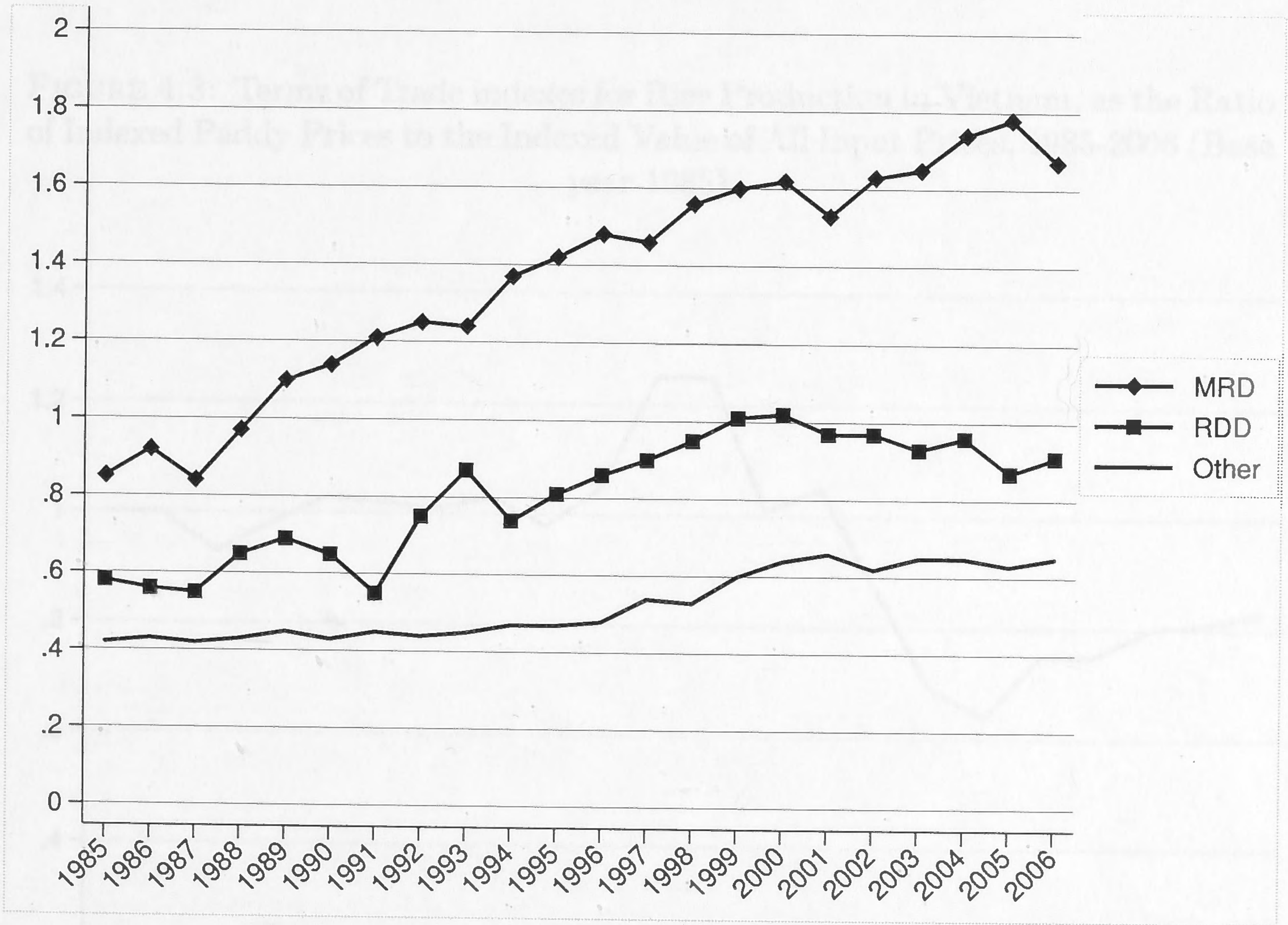


for input weighted shares v_i, for periods s and t. For convenience, results are summarized across eight regions, as officially defined in Vietnam: RRD (1), the Northeast (2), the Northwest (3); the North Central Coast (4), the South Central Coast (5), the Central Highlands (6), the Southeast (7) and the MRD (8). As mentioned, the RRD and the MRD are the major rice growing regions in the country. Region 7 is largely industrial, and regions 2, 3 and 6 are the poorest by conventional measures.

There is little doubt that the increase in rice production in Vietnam has been substantial, especially after the output share contracts period (1981–87), or under the major land and market reforms (1988–94 and forward) (Che et al., 2001). For the country as a whole, the indexed value of paddy rice output shows an average annual increase of 3.5 percent, as a fitted linear trend (see Figure 4.1). The largest increases in rice output occur during the period of trade liberalization (1988-94), and continue (virtually unabated) in the 'post-reform' period to 2006.

However, trends in TFP vary markedly between regions in the country. Figure 4.2 shows that not only is TFP higher in the MRD, but also that the growth in TFP is substantially larger in the MRD compared to the RRD and all other regions. As a fitted annual trend, the growth in TFP in the MRD is 4.42 percent, while in the RRD it is 2.25 and in all other regions 1.36 percent per year. This poor TFP performance is of special concern in the poorest regions of the country (regions 2, 3 and 6), where the average annual increase in TFP is less than 1.3 percent. In all cases, except for the MRD, there is also evidence of a 'slow down' in productivity after the year 2000. This is an added problem, again, in poor regions (which generally do not have a natural advantage in rice production, or sufficient water resources for wetland rice), but it is also a concern in the RRD, a major rice growing area, where farms remain relatively small and fragmented.

FIGURE 4.2: TFP (Total Factor Productivity) Indexes for the Mekong River Delta (MRD), the Red River Delta (RRD) and All Other Regions (Other) for Paddy Rice Production in Vietnam, 1985-2006. Average Annual Growth Rate in TFP by Fitted Linear Trend for the MRD is 4.42%, for the RRD is 2.15%, and for All Other Regions is 1.36%



With the reform process, the indexed price of rice increases throughout Vietnam, from the state-controlled low price in the 'communal period' to a partially controlled price during the output share contracts period, and beyond. During the trade liberalization period (1988–94), in particular, all controls over domestic prices were removed, and prices rose rapidly to world values, especially from

1993–96. After 1996 the domestic price of rice generally tracks world prices for rice on international markets, and fluctuates accordingly.

Input prices, on the other hand, increase more slowly (and uniformly) throughout the trade liberalization period, as the likely product of both increases in the output of rice and resulting patterns of economic development. Much of this increase is dominated by increases in the price of fertilizer (with some volatility), but farm wage rates also increase at an average annual rate of 1.44 percent.

The combination of output and input price changes, as the TOT, is summarized in Figure 4.3, showing a relatively stable trend until 1993, with improvement from 1993–96, due mostly to increases in domestic rice prices, after which it declines sharply (save 1998) until 2001. Although the TOT improves after 2001, it still remains below its starting point throughout the remainder of the series, to 2006. This highlights the importance of TFP increases to partly offset this trend, since increases in productivity will generate proportionally more revenues for given input use.

FIGURE 4.3: Terms of Trade indexes for Rice Production in Vietnam, as the Ratio of Indexed Paddy Prices to the Indexed Value of All Input Prices, 1985-2006 (Base year 1985)

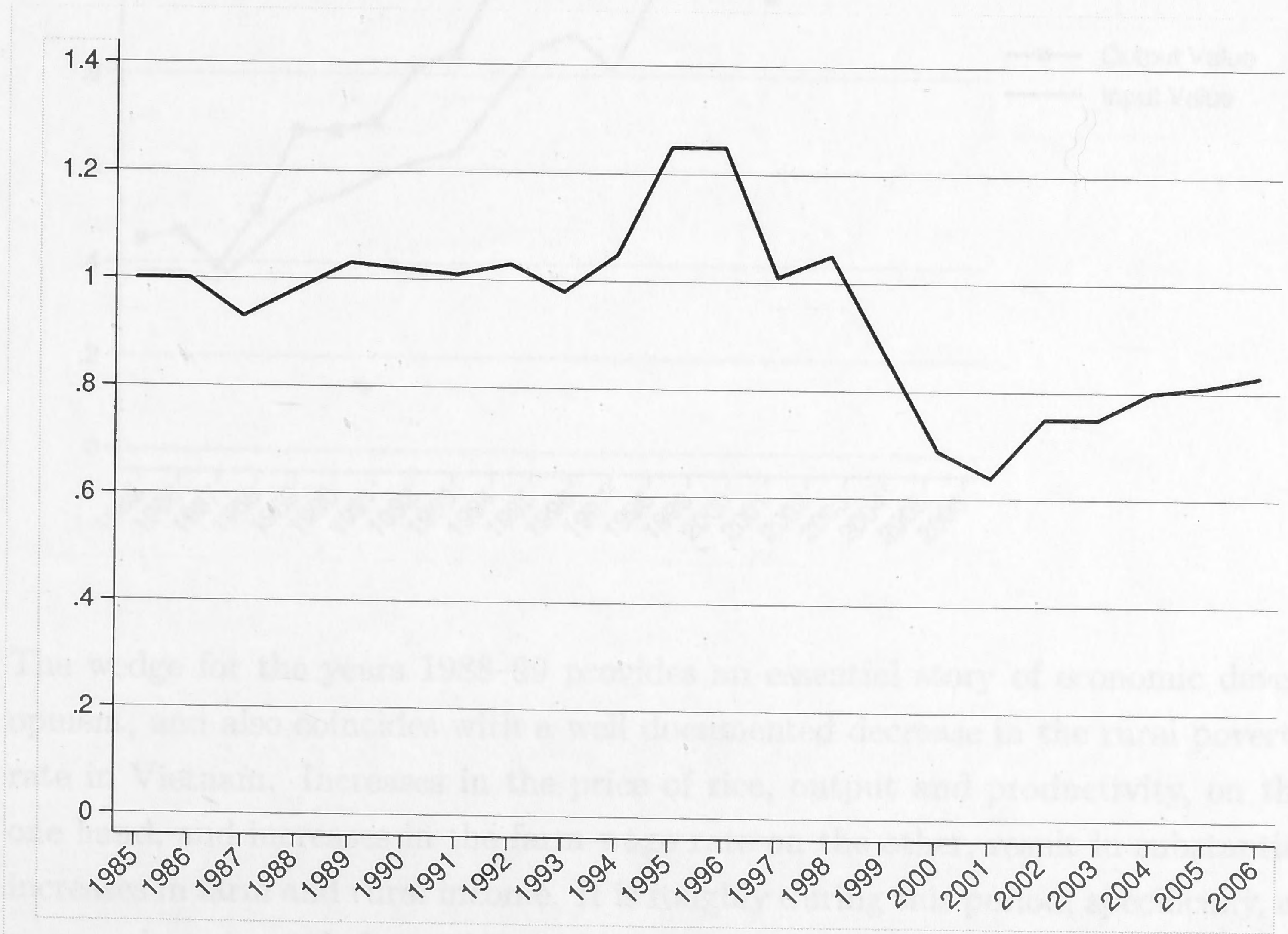
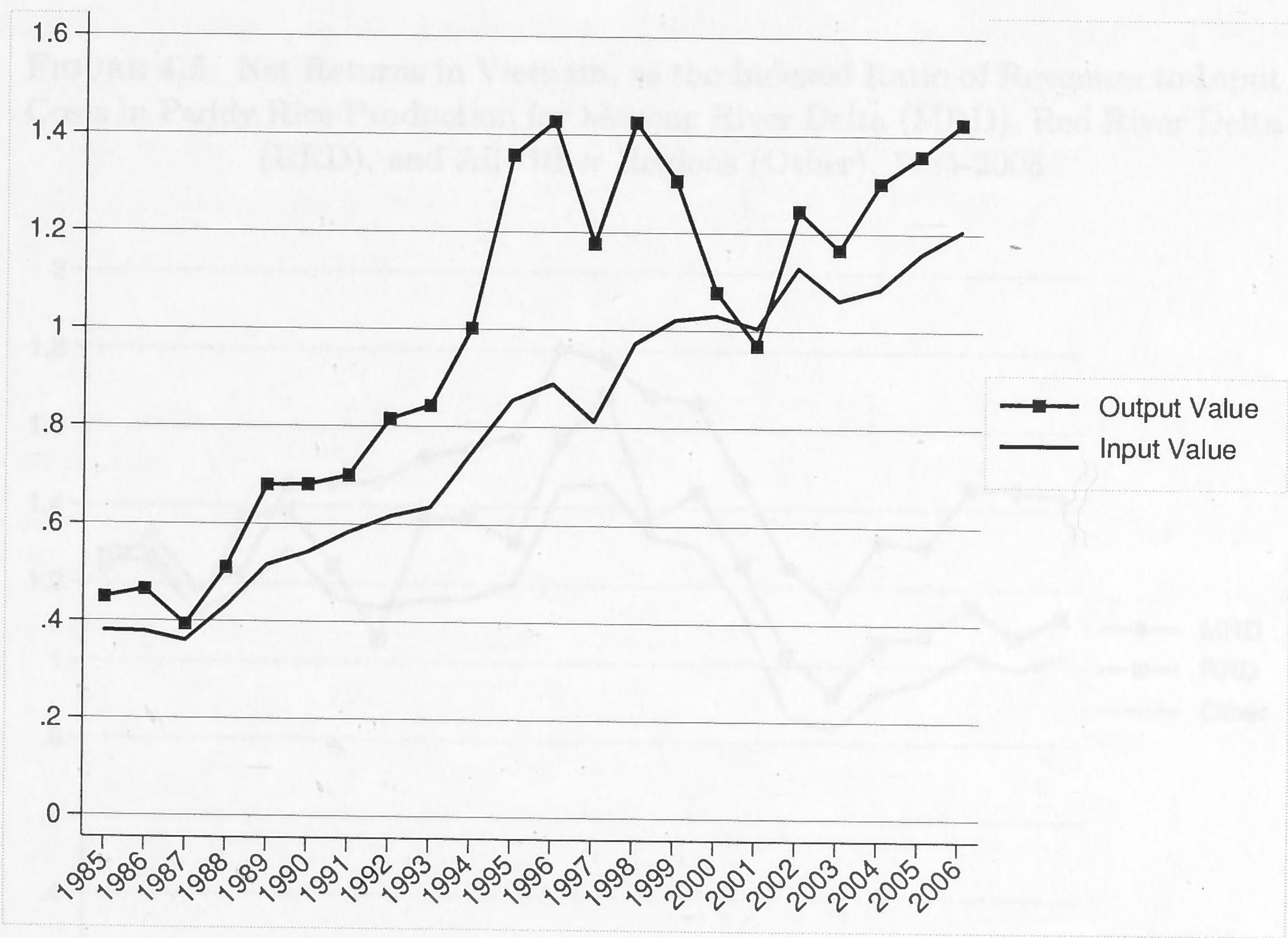


Figure 4.4 is the key graph, in effect combining all price, quantity and productivity indices together. It shows both the indexed value of paddy rice output (i.e., the indexed price multiplied by the indexed quantities of rice) and the indexed value of input expenditures (i.e., the indexed input prices multiplied by the indexed quantities of inputs). The wedge between the two lines provides a measure of 'net income' in rice production over time. For Vietnam as a whole, land and market reforms indicate substantial increases in net income from 1988–99, and especially so in the years 1993–99. In 2000–01, both the domestic and international price for rice fall dramatically, and the wedge closes.

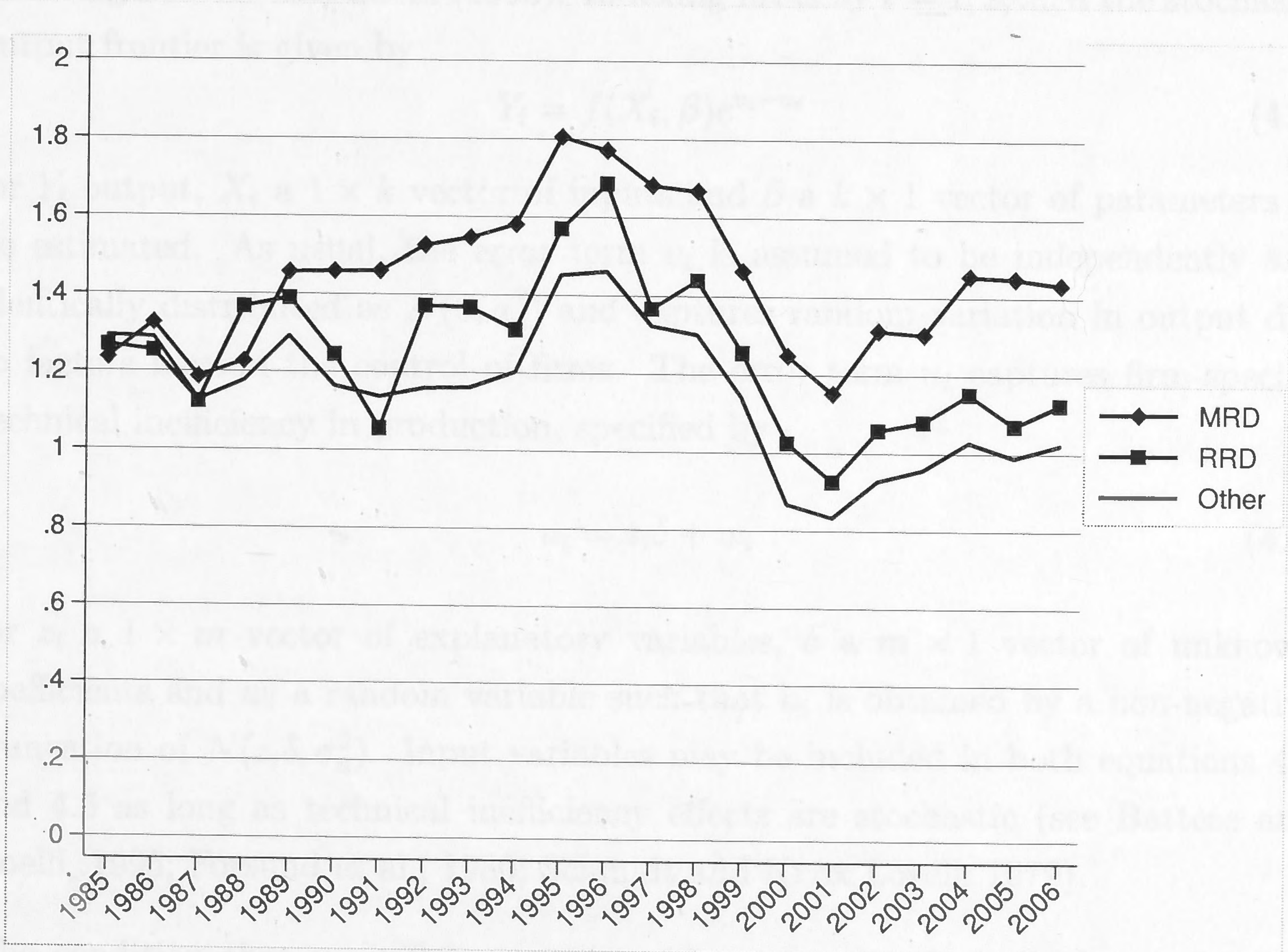
FIGURE 4.4: Net Income Measure or the Indexed Value of Paddy Rice Output Values (Indexed Output Prices Multiplied by Indexed Output Quantities) and the Indexed Value of all Input Values (Indexed Input Prices Multiplied by Indexed Input Quantities) in Rice Production in Vietnam, 1985-2006



The wedge for the years 1988–99 provides an essential story of economic development, and also coincides with a well documented decrease in the rural poverty rate in Vietnam. Increases in the price of rice, output and productivity, on the one hand, and increases in the farm wage rate on the other, result in substantial increases in farm and rural income. It is roughly during this period, specifically, as commonly measured, from 1993 to 2004, that the defined share of poor people in Vietnam "dropped by two thirds and approximately 24 million people were lifted out of poverty" (Hansen and Nguyen, 2007). Not all of this poverty reduction was due to rice production, of course, but given the large share of the population in rural areas and the predominance of rice production in rural agriculture, there is little doubt that the reform and post-reform periods had a major impact on overall living standards.[2]

Nevertheless, it is also clear that these gains are not shared equally throughout the country. Figure 4.5 shows the indexed ratio of the value of revenues to the value of input costs for paddy rice production, for selected regions, as a measure of 'net returns'. Relative to the starting point, all regions do reasonably well from 1992–98, but after 1998 both the RRD and all other regions fall (in some cases) far below the starting point. For the years 1999-2004, net returns are even less than one for areas outside of the MRD and RRD. Overall, the MRD does consistently better than all other regions with relatively large net returns, from 1989 forward.

FIGURE 4.5: Net Returns in Vietnam, as the Indexed Ratio of Revenues to Input Costs in Paddy Rice Production for Mekong River Delta (MRD), Red River Delta (RRD), and All Other Regions (Other), 1985-2006



[2]See Glewwe et al. (2002) for the distributional effects of poverty reduction in Vietnam, based on early household surveys in 1993 and 1998, and Ravallion and Van de Walle (2008) for the recent welfare impacts of land reform.

4.5 Frontiers and Efficiency

The importance of the level and changes in TFP and resulting changes in net income highlight the importance of potential efficiency gains that accompany further land and market reform. The following sections use stochastic production frontiers and inefficiency models to isolate the key constraints on efficiency gains (as a component of TFP), and what policy measures might be most suitable to increase efficiency.

4.5.1 Stochastic Frontiers and Inefficiency

Stochastic production frontiers were first developed by Aigner et al. (1977) and Meeusen and van den Broeck (1977). The specification allows for a non-negative random component in the error term to generate a measure of technical inefficiency, or the ratio of actual to expected maximum output, given inputs and the existing technology. The idea can be readily applied to both cross section and panel data, following Battese and Coelli (1995). Indexing firms by $i = 1, 2, ..., n$ the stochastic output frontier is given by

$$Y_i = f(X_i, \beta)e^{v_i - u_i} \tag{4.5}$$

for Y_i output, X_i a $1 \times k$ vector of inputs and β a $k \times 1$ vector of parameters to be estimated. As usual, the error term v_i is assumed to be independently and identically distributed as $N(0, \sigma_v^2)$ and captures random variation in output due to factors beyond the control of firms. The error term u_i captures firm-specific technical inefficiency in production, specified by

$$u_i = z_i\delta + w_i \tag{4.6}$$

for z_i a $1 \times m$ vector of explanatory variables, δ a $m \times 1$ vector of unknown coefficients and w_i a random variable such that u_i is obtained by a non-negative truncation of $N(z_i\delta, \sigma_u^2)$. Input variables may be included in both equations 4.5 and 4.6 as long as technical inefficiency effects are stochastic (see Battese and Coelli, 1995; Forsund et al., 1980; Schmidt and Knox Lovell, 1979).

The condition that $u_i \geqq 0$ in equation 4.5 guarantees that all observations lie on or beneath the stochastic production frontier. A trend can also be included in equations 4.5 and 4.6 to capture time-variant effects. Following Battese and Corra (1977) and Battese and Coelli (1993), variance terms are parameterized by replacing σ_v^2 and σ_u^2 with $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2/(\sigma_v^2 + \sigma_u^2)$. The technical efficiency of the i-th firm for the basic case can be defined as

$$TE_i = \frac{E(Y_i \mid u_i, X_i)}{E(Y_i \mid u_i = 0, X_i)} = e^{-u_i} = \exp(-z_i\delta - w_i) \tag{4.7}$$

and clearly must have a value between zero and one. The measure of technical efficiency is thus based on the conditional expectation given by equation 4.7, given the values of $v_i - u_i$ evaluated at the maximum likelihood estimates of the parameters in the model, where the expected maximum value of Y_i is conditional on $u_i = 0$ (see Battese and Coelli, 1988). Efficiency can be calculated for each individual firm per year by

$$E[\exp(u_i) \mid v_i + u_i] = \frac{1 - \Phi(\alpha_a + \gamma(v_i + u_i))/\sigma_a}{1 - \Phi(\gamma(v_i + u_i)/\sigma_a)} \exp\left[\gamma(v_i + u_i) + \sigma_a^2/2\right] \tag{4.8}$$

for $\sigma_a = \sqrt{\gamma(1-\gamma)\sigma^2}$ and $\Phi(\cdot)$ the density function of a standard normal random variable (Battese and Coelli, 1988). The value of gamma equals zero when there are no deviations in output due to inefficiency, and equals one when there are no deviations in output due to random effects, or variance in v.

4.5.2 Farm Survey Data (2004)

4.5.2.1 Econometric Specification

The first frontier estimate uses survey data obtained from a selection of 388 farms producing rice from 32 communes across 8 provinces in the RRD and MRD, with a roughly equal split of farms and communes in each area. The survey, a targeted sample of farmers primarily producing rice, was carried out by the authors from August to December 2004, with detailed collection of rice output and input data, as well as farm specific characteristics. The main areas from which farms were selected in the MRD are Soc Trang, Tra Vinh, Vinh Long and Can Tho; and, in the RRD, from Ha Tay, Nam Dinh, Thai Binh and Nam Ha. Summary statistics are provided in Table 4.1. The specification for the stochastic production frontier is given by

$$\begin{aligned} \ln Y_i = \beta_0 + \beta_1 \ln K_i + \beta_2 \ln LAB_i + \beta_3 \ln LAN_i + \beta_4 \ln F_i \\ + \beta_5 \ln P_i + \beta_6 MRD_i + v_i - u_i \end{aligned} \tag{4.9}$$

for Y the output of paddy rice in kilograms, K capital in machinery hours, as the sum of hours a farm uses tractors in land preparation and transportation, as well as hours employed for pumps and threshing machines, LAB working days, as the sum of family and hired labour, LAN total cultivated rice land in hectares times the number of crops over the year, F kilograms of fertilizer used, P pesticides in kilograms and MRD a binary variable for MRD rice farms. The inefficiency model in this case is

$$u_i = \delta_0 + \delta_1 SIZE_i + \delta_2 PLOTSIZE_i + \delta_3 SOIL_i \\ + \delta_4 IRR_i + \delta_5 ED_i + \omega_i \tag{4.10}$$

for $SIZE$ the area of rice land (both leased and directly controlled by the household) in hectares, $PLOTSIZE$ the average size of plots for rice land on a given farm, or the area of rice land in hectares, divided by the number of plots, as a proxy for land fragmentation, and $SOIL$ a measure of soil quality, ranked in decreasing order (from 1 to 6), based mainly on the chemical composition of the soil. This is an ordinal ranking, tied to land tax levies by land quality in Vietnam, and does not literally measure relative differences across land quality types (e.g., type 2 land is not twice as good as type 4 land). IRR is a measure of water availability (natural and irrigated), ranked in decreasing order (from 1 to 4), obtained by asking farmers to rank their irrigation conditions, based on the level and difficulty of supplying water and drainage. The ranking from 1 to 4 is simply given by very good, good, fair or poor. ED is the level of education of the farm decision maker, categorized by levels: primary, secondary, high school and higher education.

Additional likelihood ratio (LR) tests to confirm specification are summarized in Table 4.2. Correct critical values from a mixed χ-squared distribution (at the 1 percent level of significance) are drawn from Kodde and Palm (1986). The relevant test statistic is

$$LR = -2\{\ln[L(H_0)/L(H_1)]\} = -2\{\ln[L(H_0)] - \ln[L(H_1)]\} \tag{4.11}$$

where $L(H_0)$ and $L(H_1)$ are the values of the likelihood function under the null and alternative hypotheses, respectively. The null hypothesis that technical inefficiency effects are absent ($\gamma = \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$) and that farm-specific effects do not influence technical inefficiencies ($\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$) in equation 4.10 are both rejected, as is $\delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$. Finally, the null hypothesis that $\gamma = \sigma_u^2/(\sigma_v^2 + \sigma_u^2) = 0$, or that inefficiency effects are not

TABLE 4.1: Summary Statistics for Key Variables in Paddy Rice Production for the Farm Survey Data Set, 2004

Variables	Units	Mean	Std dev	Min	Max
Output (Y)	kg	10,321	12,484	690	105,593
Capital (K)	hours	36	37	4	272
Labour (LAB)	man days	156	91	32	584
Land (LAN)	ha	2.27	2.35	0.13	14.60
Fertilizer (F)	kg	814.08	610.05	73.00	5000.00
Pesticides (P)	kg	1.69	11.88	1.39	123.65
Mekong River Delta (MRD)	yes=1	0.50	0.51	0	1
Land (SIZE)	ha	0.92	0.99	0.06	7.30
Plot size (PLOTSIZE)	ha	0.45	0.58	0.02	4.50
Soil (SOIL)	rank	2.72	1.07	1	5
Irrigation (IRR)	rank	2.62	0.59	1	4
Education (ED)	level	2	0.64	1	4

stochastic, is rejected. All results indicate that stochastic effects and technical inefficiency matter and thus that traditional OLS estimates are not appropriate in this study. Specifications with interaction terms and non-linear effects of farm size on efficiency were also attempted, but these proved to be non-significant.

TABLE 4.2: Generalized Log Likelihood Ratio Tests and Parameter Restrictions for the Stochastic Production Frontier and Technical Inefficiency Models, Farm Survey Data, 2004

Null Hypothesis	Likelihood Ratio	$\chi^2_{0.99}$ Statistics	Decision
1. $\gamma = \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$	212.99	17.76	Reject H_0
2. $\gamma = 0$	7.56	5.41	Reject H_0
3. $\delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$	212.23	16.07	Reject H_0
4. $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$	211.75	14.33	Reject H_0

The critical values for thehypotheses are obtained from Table 1 of Kodde and Palm (1986).

4.5.2.2 Results for Farm Survey Data

Results for the farm survey data set are reported in Table 4.3. The coefficients for capital, labour, land, fertilizer and pesticide in the stochastic production frontier model are 0.15, 0.13, 0.41, 0.21 and 0.06, respectively. The value of the coefficient for the binary variable MRD illustrates the advantages of growing rice in the south, compared to the north. This value is 0.17 and is consistent with the measured difference in TFP between the RRD and the MRD, illustrated in Figure 4.2.

TABLE 4.3: Parameter Estimates of the Stochastic Production Frontier and Technical Inefficiency Models for the Farm Survey Data, 2004 (Equations [4.9] and [4.10])

	Coefficient	T–ratio
Stochastic production frontier model		
Constant	6.56(0.228)	28.83
Capital (K) (ln)	0.15(0.030)	4.89
Labour (LAB) (ln)	0.13(0.035)	3.62
Land (LAN) (ln)	0.41(0.036)	11.23
Fertilizer (F) (ln)	0.21(0.030)	6.91
Pesticide (P) (ln)	0.06(0.016)	3.88
MRD	0.17(0.037)	4.43
Technical Inefficiency Model		
Constant	0.59(0.084)	7.05
Land (SIZE)	-0.14(0.023)	-6.39
Plot size (PLOTSIZE)	-0.07 (0.023)	-2.82
Soil (SOIL)	0.05(0.008)	6.41
Irrigation (IRR)	0.05(0.012)	4.33
Education (ED)	-0.05(0.011)	-4.55
Gamma	0.67	2.26
Sigma-squared	0.01	13.86
Log-likelihood	273.25	
Mean Efficiency (%)	86	
Number of observations	388	

All coefficients are significant at the 0.01 level. Standard errors are in parentheses.

Of particular interest, however, are the inefficiency results. The coefficients on soil and irrigation are positive as expected, since the ranking on soil and irrigation is in decreasing order of quality, implying that higher quality soil and better irrigation increase efficiency. It is also clear that more educated farmers are also more efficient. The coefficients on $SIZE$ and $PLOTSIZE$ indicate the loss in efficiency from current land use practice. The estimates indicate that larger farms and farms with larger average plot size are more efficient. The latter in particular indicates a potential issue with land fragmentation, and the constraints to consolidation. Admittedly, average plot size on a given farm is a crude indicator of fragmentation, since it lacks a measure of distance between plots or whether plots are contiguous or not. However, it is also clear from the discussion in section 2 above that the smaller the average plot size on a farm the more likely it is that these plots are not contiguous. This is especially so in the north, where, as indicated, small and highly fragmented farms predominate. Frontier estimates by Van Hung et al. (2007), on a smaller survey data set for 188 farms in the north only, near Hanoi in the RRD, in the year 2000, show a comparable negative relationship between the number of plots and farm efficiency.

4.5.3 VHLSS Data (2004)

4.5.3.1 Econometric Specification

The second frontier estimate uses VHLSS data for 3,671 households largely engaged in rice production (from a total of more than 9,000 households surveyed) in 2004. The cluster survey sample applies to the entire country, not only to the main rice growing areas. Rice output of the households in this sub-sample of 3,671 farms accounts for more than 75 percent of total household annual crops in terms of quantity and more than 78 percent in terms of value. Summary statistics are listed in Table 4.4. Log-likelihood ratio tests (see Table 4.5) generate a specification for the stochastic production frontier of the form

$$\begin{aligned} \ln Y_i &= \beta_0 + \beta_1 KR_i + \beta_2 \ln LAB_i + \beta_3 \ln HLAB_i + \beta_4 \ln LAN_i \\ &\quad + \beta_5 \ln F_i + \beta_6 \ln P_i + \beta_8 MRD_i + v_i - u_i \end{aligned} \tag{4.12}$$

with an inefficiency model given by

$$\begin{aligned} u_i = \delta_0 + \delta_1 ED_i + \delta_2 FRAG_i + \delta_3 CERT_i \\ + \delta_4 SOIL_i + \delta_5 EXT_i + \delta_6 \ln CRE + \omega_i \end{aligned} \tag{4.13}$$

for Y the output of paddy rice in kilograms, produced over the twelve months prior to the survey date, KR the value of rented capital in rice production (tractors, machines, tools and implements in thousand Vietnamese Dong (VND) and LAN the amount of area in hectares that the household uses for annual crop production, regardless of its ownership. Labour comes from two sources: LAB is household labour (in hours) and $HLAB$ is hired labour in rice production (in thousand VND). Expenditures on fertilizer (F) and pesticide and herbicide (P) in rice production are all measured in thousand VND. *MRD* is a binary variable for households in the MRD.

TABLE 4.4: Summary Statistics for Key Variables in Paddy Rice Production for the VHLSS Survey Data Set, 2004

Variables	Units	Mean	Std dev	Min	Max
Paddy rice output (Y)	kg	3,878	5,931	75	120,750
Capital rented (KR)	000 VND	646	1,145	0	18,400
Labour (LAB)	hours	2,485	1,918	0	16,048
Labour hired (HLAB)	000 VND	336	1,083	0	36,000
Land (LAN)	ha	0.56	0.73	0.02	10.00
Fertilizer (F)	000 VND	1,195	1,871	0	34,000
Pesticide (P)	000 VND	388	996	0	19,800
Mekong River Delta (MRD)	yes=1	0.16	0.36	0	1
Education (ED)	years	6.62	3.69	0	17
Land Fragmentation (FRAG)	index [0,1]	0.55	0.30	0	0.99
Land with LUC (CERT)	ratio	0.80	0.40	0	1
Soil (SOIL)	rank	3.20	1.37	1	6
Access to extension services(EXT)	yes = 1	0.48	0.50	0	1
Credit (CRE)	10 mill VND	0.17	0.51	0	9

TABLE 4.5: Generalized Log Likelihood Ratio Tests and Parameter Restrictions for the Stochastic Production Frontier and Technical Inefficiency Models, VHLSS Data, 2004

Null Hypothesis	Likelihood Ratio	$\chi^2_{0.99}$ Statistics	Decision
1. $\gamma = \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$	210.42	19.38	Reject H_0
2. $\gamma = 0$	124.36	5.41	Reject H_0
3. $\delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$	154.04	16.07	Reject H_0
4. $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$	72.68	14.33	Reject H_0

The critical values for thehypotheses are obtained from Table 1 of Kodde and Palm (1986).

In the inefficiency model, *ED* is years of schooling of the household head and *FRAG* is a measure of land fragmentation given by a Simpson index, defined as

$$\left(1 - \frac{\sum_i a_i^2}{A^2}\right) \tag{4.14}$$

where a_i is the area of the i-th plot and $A = \sum_i a_i$ is total household farm size. A value of zero for the Simpson index indicates that a household has only one parcel or plot of land. The closer the index is to one, the more fragmented is household farm land, or the more numerous are the number of plots. $CERT$ designates that a farm household holds a land certificate title (measured as a ratio of land under title to total land size), allowing for the sale or lease of all or some plots of land, and $SOIL$ is a measure of soil quality, as in the farm survey data set, ranked in decreasing order (from 1 to 6). Not surprisingly, the measure of soil quality varies more and the mean is higher (meaning less quality) in the VHLSS data, with its broader coverage across all types of land, and throughout the country, compared to the farm survey data, which concentrates only on rice farms in the MRD and RRD (see tables 4.1 and 4.4). EXT is a binary variable indicating whether the farmer visited an extension services office, attended meetings to seek advice or guidance on cultivation practices or raising livestock, or was visited on the farm by an extension staff officer. In the summary data, 48 percent of those surveyed accessed these services. The value CRE is total farm household credit from banks and other financial institutions for agricultural production during the year in ten-millions of VND.

4.5.3.2 Results for VHLSS Data

Results for the VHLSS data set are reported in Table 4.6. Estimated input coefficients are comparable with the results for the farm survey data set for capital, land, fertilizer and pesticide. The binary variable MRD again indicates the advantages of growing rice in the MRD. Results indicate that better educated farmers and higher quality soil increase efficiency across farms. The estimated coefficient on $SOIL$ in the VHLSS estimates (0.48 compared to 0.05 in the farm survey data) reflects the importance of the contrast in soil quality outside of the major rice growing regions, and its effect on inefficiency.

The VHLSS estimates also show, more pointedly, the effect of land fragmentation on efficiency, using a Simpson index. The more fragmented is a household farm, or the larger the number of plots, the lower is efficiency, with a coefficient value of 0.25.

Of added interest here are coefficient estimates on land use certificate, access to extension services and credit. As mentioned, a proper land use certificate is essential not only for the ease of acquiring, selling or leasing land, but it also provides

TABLE 4.6: Parameter Estimates of the Stochastic Production Frontier and Technical Inefficiency Models for VHLSS Data Set, 2004 (Equations [4.12] and [4.13])

	Coefficient	T–ratio
Stochastic production frontier model		
Constant	6.40(0.072)	89.50
Capital rented (KR) (ln)	0.09(0.006)	17.30
Labour (LAB) (ln)	0.03(0.006)	4.79
Hired labour (HLAB) (ln)	0.02(0.003)	7.84
Land (LAN) (ln)	0.52(0.009)	56.76
Fertilizer (F) (ln)	0.16(0.006)	27.47
Pesticide (P) (ln)	0.08(0.005)	16.92
MRD	0.21(0.022)	9.80
Technical Inefficiency Model		
Constant	-2.28(0.562)	-4.06
Education (ED)	-0.07(0.015)	-4.62
Fragmentation (FRAG)	0.25(0.100)	2.48
Ratio of land with LUC (CERT)	-0.24(0.075)	-3.19
Soil (SOIL)	0.48(0.084)	5.74
Access to extension services (EXT)	-0.08(0.041)	-1.98
Credit (CRE) (ln)	-0.07(0.026)	-2.49
Gamma	0.88	42.55
Sigma-squared	0.52	6.02
Log-likelihood	-1315.25	
Mean Efficiency (%)	77.49	
Number of observations	3,671	

All coefficients are significant at the .01 level except access to extension services (EXT), which is significant at the .05 level. Standard errors are in parentheses.

the only ready source of collateral for farm loans. Those farms with a proper certificate are more efficient, as are those that have access to agricultural extension services and credit. The result on credit, showing that farms receiving more credit are more efficient, along with the fact that only 20 percent of households have access to funds, is consistent with recent findings suggesting substantial credit constraints for all rural households in Vietnam (Bao Duong and Izumida, 2002).

4.6 Conclusion

Extensive land and market reform in Vietnam has resulted in dramatic increases in rice output over the past thirty years. Results show that TFP increases considerably in the major rice growing areas (the MRD and RRD areas) during the early years of land and market reform, but with evidence of a productivity 'slow-down' since 2000 in all regions except the MRD. TFP in the MRD remains much higher than in the RRD, and TFP in other regions (excluding the RRD and especially in poor areas) remains virtually unchanged throughout the entire period. Terms of trade, net incomes and returns are also favorable throughout the reform period, providing much of the explanation for increased incomes and poverty reduction during this time, noting that overall performance has worsened considerably since 2000–01. The differences over time and by region speak directly to existing land use regulations and practices and suggest calls for further land and market reform. In this regard, additional frontier and efficiency model estimates illustrate the remaining institutional and policy constraints, including existing restrictions on land consolidation and conversion and poorly developed markets for land and capital. Estimates show that larger and less fragmented farms, and those that are better irrigated, with higher quality spoil, a clear property right or land use certificate and access to agricultural extension services and credit are more efficient.

With this in mind, it seems clear that the mandate to grow rice in every province, at least in terms of a narrowly defined efficiency criteria, is inappropriate. Productivity and efficiency are both substantially larger in the MRD and RRD areas, where rice production has a clear comparative advantage. This shows up repeatedly in both TFP and related measures, as well as in frontier and inefficiency models in terms of the magnitude of the binary variable for the MRD (and its effect on output). It is also indicated by coefficient estimates that measure the effects of irrigation and soil quality on efficiency. Land in the MRD and RRD is of much better quality in this regard, and naturally suited to rice production. Land policy (formal or in practice) which makes it difficult for land to be converted to other uses is thus difficult to justify.

The same can be said for land consolidation. If farms that are larger and less fragmented are more efficient, practical restrictions on land size need to be relaxed and a more active real estate market for land needs to be provided, encouraging low-cost and efficiency enhancing land transfers. A necessary and straightforward prerequisite for this is well-defined land use certificates, covering every parcel of land, something that Vietnam has yet not been able to accomplish. This may also partly resolve problems with credit availability, as would a significant extension of the 20 year lease provisions on parcels of agricultural land. Without a land use certificate, or with limited remaining tenure, it is difficult, if not impossible, to secure a loan, much less convert and consolidate land. The original land and market reforms, as dramatic as they were, have not gone far enough to secure property rights or provide sufficient or suitable markets for land and credit.

There are at least three issues that warrant further research. First, it would be useful to have a more refined measure of land fragmentation than either the Simpson index or average land size and number of plots used here, one that includes distance and a spatial representation of non-contiguous plots. There is partial data available for this in Vietnam, but more needs to be collected. Second, the estimates would benefit from additional measures of rural services. The only variable used here, access to agricultural extension services, as a simple binary variable, matters to efficiency, but so too must variables like rural infrastructure (e.g., roads, water rights and quarantine and surveillance measures) and specific cultivation practices, including the use of rice hybrids. Unfortunately, there is a lack of broad rice farm survey data to provide such estimates. Finally, and perhaps most importantly, there needs to be a clear investigation into the precise nature and cause of the thin or poorly developed agricultural land and credit markets in Vietnam, and what specific policies might be best to help relax these constraints.

4.7 Appendix

4.7.1 Data Sources and Adjustments

Data for TFP and related measures (1985-2006) and for the 1991-1991 balanced panel data set is drawn mainly from the Statistics Department of Agriculture, Forestry and Fisheries, 1991-2006, data sources obtained from the General Statistics Office of Vietnam, including VHLSS data, related project investigations, studies and reports by Vietnamese organizations, such as the State Planning Committee (SPC), the Ministry of Agriculture and Food Processing Industry (MAFI), the Ministry of Water Resources (MWR), the Department of Prices and Markets (DPM) (formally known as the State Department of Price (SDP)), and international organizations such as the World Bank and the Food and Agriculture Organization (FAO). The details of the structure of rice production (especially for the early data series) are extracted from the Surveys of Rice Production in the RRD and the MRD by Cantho University, funded by the International Rice Research Institute (see Nguyen, 1995; Vo, 1995).

It should be noted that from 1985-2002 there were 60 defined provinces in Vietnam based on the GSO statistics and administrative units. However, beginning in 2003, provinces were redefined into 64 provinces based on the GSO statistics and administrative units. In this study the 'new' provinces are aggregated into the previous provinces in the data set before 2003, for consistency. In particular, Can Tho, Dak Lak and Lai Chau refer to Can Tho and Hau Giang, Dak Lak and Dak Nong, and Lai Chau and Dien Bien provinces. Regions are as currently defined by the GSO. Primary data for 1985-1999 is obtained from Che et al. (2006) and Kompas (2004). The data set for 2000-02 is from Kompas (2004). In general, prices are measured in constant 2006 USD, and converted to Dong where appropriate. Data assembly and construction is as follows.

1. Output quantity and prices. Paddy output is drawn from SDAFF (2001, 2006) and General Statistic Office (2008b) under the category of 'production of paddy by province'. The time series of rice prices by province is computed from a number of sources, with recent data provided by the GSO. For the period 1985-2003 price data is based on Kompas (2004) and Che et al. (2006), most of which is obtained from the Department of Prices and Markets (DPM). A rice equivalent for output is chosen rather than rice output alone since in the same rice fields farmers usually overlap production with other short-term cereal crops, such as sweet potatoes and

maize. There are multiple crops per year in many areas. Specific time series data for rice output is from SDAFF (1991) and MAFI (1991) for the period 1976-90, from SDAFF (2001) for 1990-93, General Statistic Office (1995) for 1994 and SDAFF (2001) for 1995 to 1999. All measures were verified by alternative data sets contained in SDAFF (2001) for the years 1975-1999. Updates were obtained from SDAFF (2006).

2. Land quantity and prices. The time series for 'planted area of paddy' is obtained from SDAFF (1991, 1992, 2001, 2006), SPC (1995) and General Statistic Office (2008b). The Vietnamese government divides the soil quality of land into seven levels and levies land tax depending on quality. A study by World Bank (1994) distinguished the quality of soil into five levels in terms of cultivated area. Soil conditions and irrigation is generally much better in the RRD and MRD, compared to other regions (MWR, 1994). Land-use price variables are defined as the cost of land use, or the average tax levies per one sown hectare in terms of value. The tax levies are required to be paid to government for the right of using land, which depend on land quality (by rank from type 1 to 5). Land taxes for rice land are based on the gross value of rice production (SDAFF, 2006). It is assumed that the land price indices are coincident with the gross value of rice area (as the multiple of rice output overall crops per year and the price of rice).

3. Labour quantity and prices. Data for the quantity of labour is obtained by multiplying average man-days worked per hectare by the number of hectares in a given rice cultivation area. The rice cultivated area is obtained from SDAFF (1991, 2001, 2006) and General Statistic Office (2008b). Total labour for rice production is calculated from total rice planted (in area) and average labour used for rice production per hectare. Average man-day working requirements includes work for land preparation, transplanting, weeding and harvesting, originally based on the survey of rice production by the Cantho University (1990-1995), as detailed in Nguyen (1995) and Vo (1995). The data on the price of labour for paddy production is estimated from average labour costs by the SRP by the Cantho University, DPM (1995, 2002, 2005) and General Statistic Office (2006b), for the RRD and MRD. For 2003-06, the labour price variable is estimated using 2002 as a base year and the movement in the wage index for rice production, estimated as the average annual change in labour costs for rice cropping per hectare (SDAFF, 2006).

4. Material inputs and prices. Materials for paddy production are largely composed of rice seeds and preparation, fertilizer and insecticide (of these fertilizer is

the largest component, representing at least 30 to 40 percent of total costs DPM (1995, 2002, 2005) and General Statistic Office (2006b). For the period prior to 2002 material input quantities are partly measured in terms of a 'urea-used equivalent', or total planted paddy area multiplied by the rate of fertilizer used per hectare per rice crop. The rates of fertilizer use for paddy production per hectare per rice crop are obtained from the SRP for the RRD and MRD and SDAFF (2001, 2006). This rate is adjusted in some non-principal rice growing provinces, based on reports provided by the GSO. For the period of 2000-02, material inputs are estimated as a multiple of the growth indices of total fertilizer consumption used by Vietnam (FAO, 2007), using 1999 as a base year, and the actual current fertilizer use by provinces in 1999 (provided from Kompas, 2004).

In Kompas (2004), material inputs include the nutrition content of all fertilizers (organic and chemical), insecticides and seeds. The conversion factor used to aggregate organic and chemical fertilizers is similar to that used by Tang (1980) and Sicular (1988). The amount of organic fertilizer for the rice industry is obtained from the total amount of organic fertilizer used for agriculture. Organic fertilizer for agriculture is assumed to be supplied from two main sources: night soil and large animal manure (buffaloes, cattle and pigs). Population-adjusted night-soil is estimated based on the size of the rural population (General Statistic Office, 2008b) multiplied by a rural utilization rate (0.9). The standard number of large animals equals the sum of buffaloes, cattle and pigs (General Statistic Office, 2008b), for which the weighted ratios are 1, 1 and 0.33, respectively. Organic fertilizer for rice production is obtained by multiplying the amount of organic fertilizer for agriculture with the weighted ratio between food grain area sown to the total sown area for cultivation. The chemical fertilizer data used for rice production is derived directly by multiplying the average amounts of chemical fertilizer used in the north (165.4 kg/ha) and the south (193 kg/ha) (drawn from Cantho University (1995)) and the rice area in every province (SDAFF, 2006). The data set for insecticides is constructed by multiplying the average use of insecticide per hectare in the year 1992, or 5.8 kg and 7.6 kg in the north and south, respectively, and the total rice area (SDAFF, 2006). In a similar manner, the data for seeds are calculated from the average use of seeds per hectare, or 140 kg/ha and 240 kg/ha in the north and south, respectively, multiplied by the total rice area (General Statistic Office, 2008b). The time series for chemical fertilizer is calculated from the average amounts of chemical fertilizer used per hectare multiplied by cultivated area in each year (SDAFF, 2001, 2006). The time series data for insecticides and seeds are calculated from the average use of insecticide and seeds per hectare (SDAFF, 2001, 2006) multiplied by rice area for each year (SDAFF, 2006). To verify, an updated measure of fertilizer (in terms of quantities) for 2003-2006 is estimated from the trend of average fertilizer use in the South East Asia (FAO, 2007), using the 2002 as a base year.

5. Capital quantity and prices. The capital variable for 1985-1999 is based on Kompas (2004), following a similar approach to that used by McMillan et al. (1989), and assumes that the physical capital can be represented by the capacity of tractors in combination with a buffalo equivalent measure. The conversion from the number of draught animals to the capacity of tractors is based on well-known observations in Pakistan (see Blomqvist, 1986), indicating that a bullock-day (a pair of bullock working 8 hours) is approximately the same as a tractor-hour, with a typical tractor being between 15 and 25 horsepower. In the Vietnamese case, we assume that one cattle or buffalo-day is equivalent to roughly 0.6 bullock-days (or 14 hours of work by one pair of cattle or buffalo is roughly 8 hours of work by one bullock), with a typical tractor being 15 horsepower. The data sources for the capacity of tractors, number of buffaloes and cattle are provided from MAFI (1991, 1994); SDAFF (2001), and for recent series by SDAFF (2006) and General Statistic Office (2008b). The capital measure used for rice from 1985-1999 is drawn from (Kompas, 2004). The capital measure used from 2000-02 is estimated from the planted area and the average capital cost for rice from (DPM, 2005). The updated capital quantity variable for 2003-2006 is estimated and verified from the trend of tractors used in the South East Asia (FAO, 2007), using the 2002 as a base year. Capital prices for 1985-1999 are obtained from Che et al. (2001) and Kompas (2004), with additional details for the early part of this series provided in Che et al. (2006). An updated series is drawn from district level data obtained from General Statistic Office (2008b).

Chapter 5

Food Security and the Poor: Regional Effects of Rice Export Policy on Households in Vietnam

5.1 Introduction

Recent dramatic increases in the price of staple foods have raised concerns over food security and the vulnerability of the poor. World food prices reached their peak in the second quarter of 2008, with wheat and maize three times and rice five times more expensive than at the beginning of 2003 (Von Braun, 2008). Several exporting countries with food security concerns responded to these price increases by imposing export controls and (in some cases) total bans, which further fuelled world food price increases (Childs and Kiawu, 2009; Timmer, 2008; Headey and Fan, 2008). Along with a desire to ensure domestic supplies, the export controls were often rationalised by an expressed need to protect the poor from increases in domestic food prices, since staple foods account for a large proportion of their consumption bundle. Further tensions occurred in exporting countries, many of which are developing economies, since changes in the price of staple foods differentially impact rural and urban households, with relatively poor food producers in rural areas benefiting from higher world and domestic food prices at the expense of urban households.

We focus on Vietnam's rice production and trade as a case study. Vietnam is the world's second largest rice exporter, with exports of six million tons, equivalent to 16 percent of the world trade volume in rice (Shigetomi et al., 2011). Its export revenue is approximately three billion USD, contributing roughly three percent to Vietnam's GDP (World Bank, 2009c). Vietnam is also a poor developing country with about 15 percent of the population or twelve million people living below the poverty line. Although a relatively small component of GDP, trade and trade policy in the rice sector is very important for the Vietnamese because as much as 66 percent of rural households and 77 percent of the poorest quintile in Vietnam are rice producers. Rice is also the dominant staple food in Vietnam, representing 33 recent of the total household expenditure among the poorest quintile households.[1]

Given the importance of rice, the Government of Vietnam has maintained strict control over rice exports. To do so, the Government sets an annual rice export target, which can be adjusted throughout the year, subject to changes in domestic supply and demand. The Government, at its discretion, can also suspend rice exports whenever it is deemed necessary. This control is further underpinned by the almost monopoly-like power possessed by state-owned enterprises (SOE) in the rice export market, as well as their heavy involvement in rice export policy and management. Furthermore, SOEs dominate inter-regional trade in the domestic market, especially under the national food security framework, which transfers rice from surplus to deficit regions. Given this control, geographically partitioned by two national SOEs, Vietnam's domestic rice supplies characteristically lack integration between markets across the north and the south of the country (Minot and Goletti, 1998; Baulch et al., 2008; Luu, 2003).

In 2008, in the face of rising world prices for rice, Vietnam imposed an export ban from March 25th to the end of June over concerns for food security and a desire to stabilise the domestic price of rice. This ban, together with one by India earlier on, the world's third largest rice exporter, coupled with panic purchases by rice importers, especially from the Philippines, contributed to pushing the world rice price to its peak in May of 2008 (Timmer, 2008).

This paper analyses Vietnam's rice export policy in the context of rising world rice prices. In particular, we investigate both national and sub-national impacts as well as the distributional and welfare implications of different policy scenarios. To do so, we bring together insights from a regionally-disaggregated or 'bottom-up' computable general equilibrium (CGE) model and a micro-simulation on household data. At the economy-wide level, our bottom-up CGE model is a combination of eight interacting CGE models, representing eight regions in Vietnam. To this end, it allows both national and sub-national assessments of a change in the world price

[1] Calculated using Vietnam Household Living Standard Survey 2006 (VHLSS 2006) data.

of rice on GDP, domestic prices and employment under different policy scenarios. Sub-national changes in domestic producer and consumer prices of both food and non-food items, as well as changes in factor prices and including wages for both skilled and unskilled labour generated from the bottom-up CGE model, are then inputed into household data for a further disaggregated analysis of different policy options at the household level.

We consider three policy scenarios in this paper. The first is where Vietnam maintains the status-quo with a rice export control designed to mimic the imposed export ban in 2008, and the prevailing and market-segmenting powers of the SOEs in domestic rice and export markets. In the second scenario, Vietnam still controls rice export quantities, but liberalises the rice export market domestically — a WTO commitment Vietnam has promised to deliver on since 2011. In the last scenario, we assume that Vietnam has a relatively free rice export policy, with no export controls or bans, and a nearly competitive domestic rice market.

Using a Vietnam interregional input-output table for 2005 (VIRIO 2005) for the CGE model and household survey data (VHLSS 2006) for the micro-simulation, we draw three main conclusions. First, for Vietnam as a whole, although the range of export policy options has little impact on GDP, there are substantial distributional impacts across regions and households from different export polices. Second, both rural and urban households, including poor households, benefit from free trade, even though domestic rice prices are higher. Finally, under free trade, compared to other export policies, relatively large gains accrue to rural households, where poverty is most pervasive in Vietnam, with peak gains among the middle-income cohort and a relatively symmetric distribution of gains around this group.

Despite numerous recent modelling exercises done on Vietnam's international trade integration, this research, to the best of our knowledge, is the first that uses a regionally-disaggregated or bottom-up CGE model. With its disaggregation, the model is able to capture regional dimensions of the Vietnamese economy under different market conditions. The latter is important since fully integrated market conditions are not a realistic assumption for the domestic rice market in Vietnam, making an aggregate or national CGE model for the country inappropriate.[2]

Our results also differ substantially from the literature in the sense that they capture the distributional impacts from different export polices, across the regions and

[2]See World Bank (2005) and Abbott et al. (2009, 2007) for critical reviews of modelling exercises quantifying the impact of Vietnam's global integration and Baulch et al. (2008) for fragmentation in the Vietnamese domestic rice market.

for different households. The existing literature, for example, using mostly partial equilibrium methods, shows that rice export liberalisation in Vietnam would increase food prices and average real income, making urban households worse off while rural households would be better off (Minot and Goletti, 1998). Other studies on the impact of higher food prices in Vietnam generally collaborate these findings (Ivanic and Martin, 2008; Vu and Glewwe, 2011; Phung and Waibel, 2010).[3] Although these studies provide important insights, they are done in isolation from economy-wide impacts, thus ignoring the connections between a change in the price of rice and changes in the price of inputs and non-food commodities.[4] The connection between the price of rice and the wage rate, in particular, is important for understanding the effects on urban welfare, or for those who do not necessarily grow rice, from different export policies.

5.2 The rice market in Vietnam

Vietnam has made remarkable progress in rice production over the last thirty years, moving from a large importer of rice during the period 1976–80, to now the world's second largest rice exporter in the world. About 8.5 million hectares of rice planted area, equivalent to more than 4 million ha of land, produces approximately 43 million tons of rice per year in Vietnam (General Statistic Office, 2009). More than 50 percent of the rice output is produced in the Mekong River Delta (MRD), and more than 90 percent of exported rice comes from this region (Government of Vietnam, 2008). For our purposes, there are three special aspects of rice production in Vietnam worth highlighting: (1) the Vietnamese government's control of export quantities and the role of SOEs; (2) the lack of integration between rice markets in the north and south; and (3) the details of how the Vietnamese government responded to the food crisis of 2008.

[3]The impact of higher food prices in other lower-income countries has been studied widely, with a variety of conclusions. For example, see Deaton (1989) for Thailand; Budd (1993) for Cote d'Ivoire, Barrett and Dorosh (1996) for Madagascar; and Friedman and Levinsohn (2002), Warr (2005) and Ravallion and Van de Walle (1991) for Indonesia.

[4]Ivanic and Martin (2008) is the only study on the impact of higher food prices in Vietnam which takes into account changes in wages, but only for unskilled labour.

5.2.1 Quantity controls and market power in Vietnam's rice export market

Vietnam has declared three objectives in its management of rice exports: the profitability of farmers, with attempts to guarantee a minimum return over costs, food security or 'adequate domestic supplies under any circumstances', and stable domestic prices (Government of Vietnam, 2008). A recent decree by the Prime Minister replaced the food security objective by one of 'implementing international trade commitments and ensuring efficient export supplies' (Government of Vietnam, 2010).

One of Vietnam's key measures used to achieve its objectives is to control the quantity of rice exports. Since 1992, three years after Vietnam began exporting rice, the Government has controlled rice exports by setting annual rice export targets. This target is set in consultation with Ministry of Agriculture and Rural Development (MARD), the Ministry of Industry and Trade (MIT) and the Vietnam Food Association (VFA).[5] It is based on estimates of domestic supply and demand. As a result, within a given year, the targeted annual export volume can, in principal, vary subject to changes in domestic conditions, although in practice the target and the policy surrounding it is often binding and restrictive. Evidence suggests that the policy results in both rice production and exports being below their optimal levels (Nielsen, 2003).

Export quantity controls were initially carried out through an export licensing system. At one point, SOEs had a complete legal monopoly over rice exports, with each of a limited group of 15 to 40 SOEs granted a quota that specified the amount of rice it could export (Minot and Goletti, 2000). In 1998, reforms allowed for some private and foreign-shared companies to engage in rice exports, followed by a simplification of the approval system for export businesses, which was in turn replaced by the current registration system. On May 1, 2001, the export quota system was formally abolished with the view to promoting competition among rice exporters in expanding their share in the world market.

Despite abolishing the export quota system, the government's overall control of the total quantity of rice exports has remained virtually unchanged. At its discretion, the Vietnamese government can suspend or limit rice exports whenever it is

[5]VAF is a trade organisation which consists of the SOEs and private companies that engage in rice exports. Before 2011, foreign-owned companies had to establish joint ventures with SOEs in order to export rice. These joint ventures have been considered as 'associates' but not full members of VFA.

deemed necessary, and even without export bans being imposed, no further rice export contracts can be implemented whenever the total quantity of contracted rice exports reaches the government's annual target.

In addition, there has also been little diminution of the market power of rice exporting SOEs in the face of reforms. Rice exporting SOEs are dominated by two national companies: the Vietnam Northern Food Corporation (usually referred to as 'Vinafood 1'), based in Hanoi, and the Vietnam Southern Food Corporation ('Vinafood 2'), based in Ho Chi Minh City, along with a few provincial SOEs. Vinafood 1 and Vinafood 2 were established in 1995 to strengthen the state capability of food market control and provide an instrument for domestic price stabilisation (Dang and Tran, 2008).

The overall dominance of SOEs in Vietnamese rice export market is largely explained by direct government support. First, food SOEs, such as Vinafood 1 and Vinafood 2, have access to credit from government financial institutions and promotional funds for exports at highly subsidised interest rates, often set at zero (Dang Kim Son and Tran Cong Thang, 2008, Luu, 2002). Second, SOEs with export contracts can borrow up to 100% of the value of the contract, while private companies can borrow only a maximum of 70% of the collateral that is provided in the form of their buildings and equipment (Dang and Tran). Third, SOEs are often exempted from tax and receive government assistance to facilitate purchases of rice (Dang and Tran). Finally, the two national SOEs, Vinafood 1 and Vinfood 2, are often assigned exclusive roles to execute government-to-government rice deals, generating considerable clout in overseas markets (Tsukada, 2011; Dang and Tran, 2008).[6]

Another key measure of the government to achieve its rice market objectives, especially in terms of ensuring a reasonable profit for farmers, is to set a "floor price" for rice exports. This floor price serves as the basis for negotiation between rice exporters and foreign importers. As a result, the domestic rice market price, especially in the MRD, is more or less conditioned by this floor price (Luu, 2002). Until recently, the floor price was set by the Ministry of Finance (MOF) based on recommendations from MIT, MARD, Vinafood 1 and Vinafood 2, and the VFA.

[6]The efficacy of monopoly-like power over food supplies exercised by SOEs generates differing views. Ghosh and Whalley (2004) suggest that the price control of those parastatals increases welfare. On the other hand, Dang and Tran (2008) argue that food SOEs are inefficient and they fail to deliver their objectives as well as incur large fiscal losses. For example, total net losses of the central SOEs in agriculture were estimated to be US$22 million in 1998 (Dang and Tran, 2008). Recently, there are increasing signs of deterioration in the general financial health of the SOEs (World Bank, 2011).

Since 2011, the floor price has been set by the VFA directly, based on guidelines promulgated by the MOF. It is not uncommon to observe executive staff members from Vinafood 1 and Vinafood 2 taking lead positions in the VFA. This, combined with no representation from farmers, has raised conflict of interest questions for the VFA in formulating producer prices (e.g., Phap Luat, 2010).

5.2.2 Lack of integration between domestic rice markets in the north and the south

Despite extensive market liberalisation in agricultural production in Vietnam after *Doi Moi*, domestic market integration in Vietnam has lagged considerably. This is partly explained by substantial constraints to transportation generated by geographical conditions associated with an elongated country, coupled with poor infrastructure due to long-lasting wars in the middle of the last century. Bureaucratic rigidities before 1997, where the procedures to buy and transport rice from the south to the north resembled those for trade with another country, also created considerable market segmentation[7] (Minot and Goletti, 2000).

Recent evidence suggests that the poor integration between markets in the north and the south continues (Baulch et al., 2008; Minot and Goletti, 2000), whereas markets within a region seem highly integrated (Baulch et al., 2008; Luu, 2003). This is largely explained by the position and power of the SOEs. Long distance trade tends to be dominated by the SOEs simply because they are well-resourced, supported by the government and, under the framework of the national food security policy, they are directly tasked with transferring rice from surplus to deficit regions, albeit under often market-distorted pricing. Only a few large private traders, miller-polishers and polishers can compete with SOEs in inter-regional trade. Given the small number of players and the reported inefficiency of SOEs, improvement in the north-south market integration is unlikely. By contrast, operating in markets within a region is seen as a distinct advantage for private traders given their local knowledge. In these markets, competition and the large number of participants often results in efficient outcomes and little remaining opportunity for arbitrage (Luu, 2003).

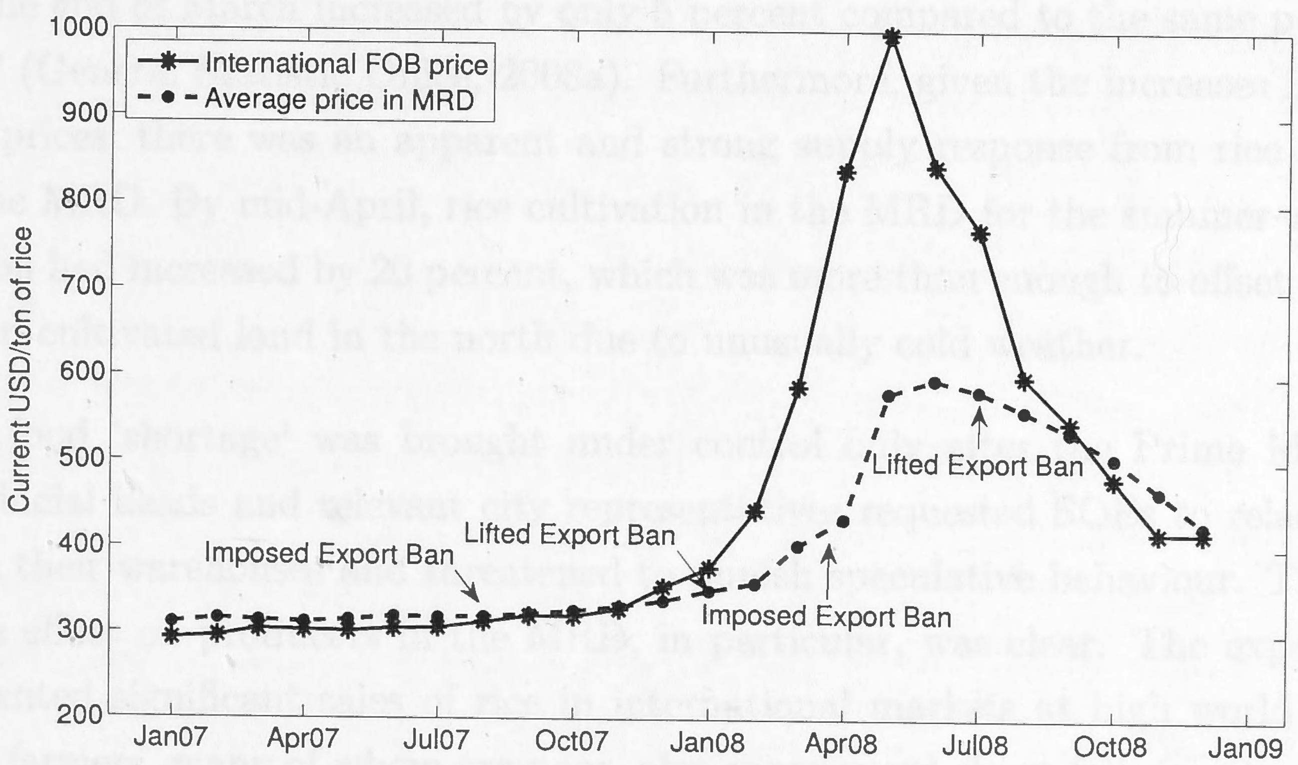
[7]In March, 1997, Decree No. 140/TTg was issued to lift internal trade restrictions on rice, to help eliminate some licenses and controls on transports.

5.2.3 The 2007–2008 food crisis in Vietnam

Recent spikes in the price of rice have generated two official responses from the Government of Vietnam. The first was a recommendation by the VFA in July of 2007 for a ban on the signing of new export contracts beyond the annual export target, effectively imposing a binding and upper-limit on exports. The government gave official approval for this action in September, 2007. This ban was removed in January 2008.

The second and more dramatic action occurred in 2008, as indicted above, when the government imposed an export ban from the 25th March until the end of June, during the peak of the global food crisis, when international prices for rice rose rapidly from 400 USD in January to roughly 1000 USD per ton in May (Figure 5.1). The ban was rationalised on the grounds of maintaining domestic food security and the control of domestic prices — with the latter objective, in large part, designed to protect the poor and urban consumers.

FIGURE 5.1: Monthly International Free-On-Board (FOB) and MRD Prices of Rice



Data on retail rice prices of MRD from General Statistics Office; data on FOB rice price of Vietnam from FAOSTAT Database, 2008.

It is clear from the evidence that these objectives were not achieved. In terms of food security, and due partly to panic hoarding by consumers and speculative delays in sales by rice wholesalers, domestic supplies of rice in stores effectively disappeared throughout most of the country[8]. For example, in late April, many retail shops were closed throughout provinces in the MRD. For those stores that remained open, both here and in the northern cities in particular, rice prices increased by the hour and many stores sold out of rice completely, or sold only in limited quantities (e.g., limits of 10 kg per customer in Ho Chi Minh City were common) (Tuoi Tre, 2008). Domestic prices were also not stabilised. Across the country, prices of staple foods increased by 6.1 and 22.19 percent monthly, as compared to 2.2 and 2.28 percent for non-staple foods in April and May (General Statistic Office, 2008b).

The effective shortage of rice was at odds with the announcements of policy makers as well as available information on domestic supplies. For example, in April, the chairman of the VFA, who was also the General Director of Vinafood 2, was quoted as saying that a "rice shortage was impossible due to the recent large harvest in the MRD and a much lower export target allowed for in international contract negotiations as compared with the previous year" (Tuoi Tre, 2008). Indeed, total production for the winter–spring season in the MRD (before the export ban) increased by 4 percent, or by 300 thousand tons, compared with harvest over the same period in the year before. On the other hand, the total rice export volume by the end of March increased by only 5 percent compared to the same period in 2007 (General Statistic Office, 2008a). Furthermore, given the increases in world rice prices, there was an apparent and strong supply response from rice farmers in the MRD. By mid-April, rice cultivation in the MRD for the summer–autumn season had increased by 20 percent, which was more than enough to offset a slight fall in cultivated land in the north due to unusually cold weather.

The food 'shortage' was brought under control only after the Prime Minister, provincial heads and relevant city representatives requested SOEs to release rice from their warehouses and threatened to punish speculative behaviour. The ultimate effect on producers in the MRD, in particular, was clear. The export ban prevented significant sales of rice in international markets at high world prices. Rice farmers, many of whom are poor, also experienced sharp falls in returns over costs, from an estimated 85 percent for their winter–spring season to only 20 percent for their summer–autumn harvest (Government of Vietnam, 2008). Indeed, toward the end of 2008, the government had to support SOEs to guarantee returns to farmers with additional rice purchases and subsidies due to the sharp fall in international demand and substantial excess domestic supplies.

[8]Much of volatility in rice prices throughout the world was attributed to hoarding behaviour (Timmer, 2010, 2012)

5.3 Method, Model and Simulation

The following sections describe the different policy scenarios used in the modelling, the bottom-up CGE model itself and the measurement of distributional impacts and changes in household welfare in the micro-simulation model.

5.3.1 Policy scenarios

Our goal is to examine the impact of a 30 percent increase in the world price of rice on Vietnam's economy and households in order to mimic (often dramatic) changes in world rice prices. The overall price change is similar to the change in world rice prices from 2005 to 2007, but less than the price-spike that occurred in mid-2008 (Croser et al., 2010; Ivanic and Martin, 2008). The percentage change in rice prices can easily be scaled in the model to generate contrast and magnified effects. We use a bottom-up CGE model to analyse the performance of the national and regional economies in terms of GDP, employment and domestic output prices, wages and other input prices. The changes in domestic output prices, wages and input prices by region, including the effects on rural and urban households as well as on producers and consumers, generated by the bottom-up CGE model, are then simulated using household data to measure the regional impacts on household welfare.

We consider three policy scenarios. The first, termed 'Quota Monopolist', resembles Vietnam's current situation where both rice export controls and the market power of SOEs are in place. The model under this scenario is designed with a (variable) administrative rice export limit and a producer tax of 15 percent to mimic the super-profits of food and rice export SOEs with an export ban. In the second scenario, designated simply as 'Quota', Vietnam still controls rice export quantities, but liberalises the rice export market — a WTO accession commitment, as mentioned, Vietnam has promised to deliver since 2011. The model under Quota is similar in design to Quota Monopolist but it does not have a producer tax and the rice market is otherwise competitive so that rice can shift (although not completely freely) throughout the country in response to price signals and excess supplies and demands. Finally, under the last policy scenario, 'Free Trade', Vietnam allows free exports, without control, and more nearly competitive domestic markets for rice. In all scenarios, prevailing fragmentation of markets between the North and the South is taken into account for by the regional dimension of our regional CGE model.

5.3.2 The bottom-up CGE model

5.3.2.1 Overview

The bottom-up CGE model used in this paper is based on the ORANI model for the Australian Economy (Horridge, 2003). To generate bottom-up and multi-regional characteristics, the top-down regional extension in the ORANI model has been replaced with a fully bottom-up regional model. In basic terms, our bottom-up CGE model can be viewed as a combination of eight interacting ORANI models, representing eight regions in Vietnam.

We use VIRIO 2005 to construct the regional characteristics of the CGE model. VIRIO 2005 is a database that covers eight regions (denoted R), representing the Red River Delta (RRD), the North East (NE), the North West (NW), the North Central Coast (NCC), the South Central Coast (SCC), the Central Highlands (CH), the South East (SE), and the MRD. The RRD and especially the MRD are the major rice growing regions, although rice is grown in almost every province of Vietnam. The SE is largely industrial, the CH is dominated by coffee production and other industrial crops, and the NCC and SCC are clearly coastal areas. The poorest regions are the NW and CH where many of the ethnic minorities live (Nguyen et al., 2012). The terrain in these regions is hilly and often mountainous and far less suitable for wet rice production. Details on specific provinces included in each region are provided in Appendix 5.6.1. VIRIO 2005 has 28 industries (I) which produce 28 commodities (C), namely: paddy, other crops, livestock and poultry, forestry, fish farming, fisheries, oil and gas, mining, processed seafood, processed rice, other agricultural processing, textiles, paper, wood, rubber, non-metallic mineral products, transport equipment, metal products, other manufacturing, electricity and water, construction, transport (margins), communication, trade (margins), financial services, public administration, hotels and restaurants, and other services.

While our bottom-up CGE model has features similar to a typical ORANI model, it differs in at least three important ways to incorporate regional features. First, regional indices are added to every variable and coefficient, increasing the dimension of the model considerably. Second, whereas the basic flow of goods and/or services in a typical ORANI database maps from two sources/destinations (i.e., the domestic economy and the rest of the world), in our bottom-up CGE model, the mapping is from R+1 sources/destinations to include all R regions and the rest of the world. Similar regional identifiers are applied to the usual ORANI structure of designated 'Margins', 'Taxes','Labour', 'Capital', 'Land', 'Other Costs' and a 'Production Tax'. Finally, instead of having a CxI ORANI dimension, the 'make matrix' in our model is a $CxRxI$ matrix to capture regional features. For simplicity, we assume that every industry is 'local' in the sense that it can produce goods and services only within its region.

5.3.2.2 Model description

The model itself consists of four agents: the household (urban and rural), the government, the investor and the foreign sector (exports and imports). It has five blocks: production, demand, market clearing, price linkages and miscellaneous blocks.

The production block in each region is made up of a set of Constant Elasticity of Transformation (CET) and Leontief production functions. Apart from the differences mentioned in the previous sub-section, the production structure is the same as that of a typical ORANI model. That is, composite intermediate commodities, primary factors and other costs are combined in fixed proportions (i.e., Leontief functions) into RxI output (activity) levels.

The demand block comprises demands for productive factors (skilled and unskilled labour, capital, and land) and demands for commodities (intermediate, household, investment, export, government, inventory, and margin demands). The demand for productive factors is given by a set of Constant Elasticity of Substitution (CES) demand functions, as are the intermediate demands for each commodity. This CES functional form implies that commodities demanded can be substituted for one another depending on their prices and the elasticities of substitution between them. Given constant returns to scale, which characterises the model's production technology, the competitive 'Zero Pure Profits' condition is imposed to equate output price to its marginal cost of production. In brief, each of the C commodities at the base of the regional production structure is generated by using the commodities bought from the other R regions and abroad with a CES and Leontief technology. These outputs are then transformed into C goods and sold in R regions and abroad based on the CET function, which is employed to model trade flows among regions and the rest of the world.

Household demand is a combination of a Stone-Geary and a CES function. Instead of using the typical ORANI model's Klein-Rubin utility function, we follow Dixon and Rimmer (2005) in employing a Stone-Geary function, reflecting the assumption that households always consume a basic subsistence bundle regardless of their budget and the prices of the bundle. Any expenditure over and above this follows a standard CES functional form. At the base of the household demand structure, each of the 28 composite goods are created by goods bought from each R region or abroad with CES technology, and then combined into final consumption goods for households in each region by the Stone-Geary utility function transformation. Parameter values for the household demand are drawn from the Vietnamese Monash (VIPAG) model, which also uses the Stone-Geary utility function.

Investment demand is constructed by a combination of Leontief and CES functions. Each of the 28 composite goods is created by goods bought from each R region and abroad with CES technology, in a manner similar to household demands. They are then combined into capital for I industries in regions R using Leontief production functions.

Export demand for each region and the world is assumed to have the following specification:

$$Export(C,R) = QF(C,R)\left[\frac{P(C,R)}{e * PF(C,R)}\right]^{EXP\_ELAST(C,R)} \tag{5.1}$$

where for each good C in region R, $Export$ is real export volume; QF and PF are quantity and price shift parameters; P is the export price; $EXP\_ELAST(C,R)$ is the export demand elasticity; and e is the exchange rate. For the export demand schedule to be downward slopping, $EXP\_ELAST(C,R)$ must be negative in the model. While equation (1) looks similar to a demand equation in a typical ORANI model, here there are CxR export demand equations with regional indices R being incorporated to model trading flows between each of the eight regions and the rest of the world. The flow of domestic goods among eight regions, as usual, follows typical demand, supply and market clearing conditions.

Other demand components maintain their corresponding ORANI setup, except for basic regional designations. Government expenditure is tied to private consumption. Inventory demands for each region depend on its production volume or its imports. Margins incorporate transportation and trade services, where margin demands depend on commodity flows and are linked to intermediate, investment, private, and government demands. There are no margins for inventory demand. Apart from adding the usual regional index to all the original ORANI equations, we also assume that the region that uses a margin is also the one that delivers the margin, reflecting the fact that trade services are largely local.

The market clearing block has standard equations to ensure market clearing conditions in each market, but with regional balance. For example, in the commodity market, the usual condition that commodities produced in each region are equal to their demand is strengthened to ensure that the production of any good in each region must be equal to its use in that region and in all other regions. Likewise, the total imports of any good must be equal to its use in all regions combined.

In the labour market, aggregate regional labor supply is assumed to be fixed. However, labour is allowed to be mobile across industries so that output in industries can vary subject to price changes. Returns to labour, or wages, are indexed to the CPI to reflect short-run conditions in the labour market. As labour is mobile within a region, the real wage for each industry adjusts according to the difference between the supply and demand for labour. In the capital market, on the other hand, capital is assumed fixed across industries to concentrate on short run effects of a change in rice prices.

The price-linkage block maintains the link between the producer and the consumer prices. The gap between the two prices is taxes, by definition, which include excise, value-added taxes, duties, and margins, which include wholesale and retail charges and transportation. Finally, the miscellaneous equations block includes reporting and equations for recursive dynamics simulations, which are not needed in our study. Interested readers are encouraged to consult (Horridge, 2003) for further details on all of these equations.

5.3.3 Measurement of household welfare impacts

To measure the change in household welfare, we use a method based on Deaton (1989) as implemented in Minot and Goletti (1998). Since a household can be a consumer, or a producer, or both, its net welfare change is a combination of both consumer and producer surpluses. For an individual household, the change in consumer surplus (ΔCS) associated with the change in the consumer price of a good is simply approximated as:

$$\Delta CS \cong -q_1^d(p_2^d - p_1^d) = -q_1^d(\Delta p^d) \tag{5.2}$$

where q_1^d is the quantity demanded before the price change, p_1^d and p_2^d are the consumer prices before and after the change, and Δp^d refers to the change in the consumer price. This first order approximation reflects only the immediate impact of the price change as it does not take into account a consumer's response. The consumer's response is included in the second-order approximation, given by:

$$\begin{aligned} \Delta CS &\cong -0.5(q_1^d + q_2^d)(p_2^d - p_1^d) = -q_1^d(\Delta p^d) - 0.5(\Delta p^d)(\Delta q^d) \\ &= -q_1^d p_1^d \frac{\Delta p^d}{p_1^d} - 0.5\varepsilon_d q_1^d p_1^d \left[\frac{\Delta p^d}{p_1^d}\right]^2 \end{aligned} \tag{5.3}$$

where q_2^d is the quantity demanded after the price change, Δq^d refers to the change in consumption, and ε_d is the price elasticity of demand.

Likewise, the second order approximation for a change in producer surplus (ΔPS) associated with the change in the producer price of a good is:

$$\Delta PS \cong 0.5(q_1^s + q_2^s)(p_2^s - p_1^s) = q_1^s p_1^s \frac{\Delta p^s}{p_1^s} + 0.5\varepsilon_s q_1^s p_1^s \left[\frac{\Delta p^s}{p_1^s}\right]^2 \tag{5.4}$$

where $q_1^s$and q_2^s are quantities supplied before and after the price change, p_1^s and p_2^s are producer prices before and after the change, and Δp^s is the change in the producer price.

Since consumers and producers will generally respond to changes in prices, we mainly use the second order approximation of ΔCS and ΔPS. For calibration, we employ the recent estimate of demand elasticity for rice in Vietnam by Nguyen et al. (2009) and the average of supply elasticities for rice in the North and South of Vietnam by Khiem and Pingali (1995).[9] For other commodities, we apply estimates from various studies on Vietnam and other countries (see Appendices 5.6.2 and 5.6.3 for details). Finally, we define the sum of ΔCS and ΔPS to give a measure of net benefit (NB):

$$NB = \sum_{i=1}^{N}(\Delta PS_i + \Delta CS_i) \tag{5.5}$$

where N is the number of goods a household consumes and/or produces. To analyse distributive effects, we focus on the ratio of household net benefit to its

[9]There might be a concern that demand and supply elasticities for rice may vary with household living standards and across regions. To address these two concerns, we checked the sensitivity of our results by using alternative estimates for demand and supply elasticities by regions and quintiles in Vietnam from Minot and Goletti (2000) and Le (2008). Overall results changed only slightly and are available from the authors on request.

expenditure, or:

$$NBR = \sum_{i=1}^{N} \frac{(\Delta PS_i + \Delta CS_i)}{Y} \tag{5.6}$$

where Y is total household expenditure before the price change, and NBR is the net benefit ratio.

As indicated above, we use the Vietnam Household Living Standard Survey in 2006 (VHLSS 2006), carried out by the General Statistics Office (GSO) in Vietnam, for the micro-simulation. VHLSS 2006 is the latest household survey available before the food crisis and the various rice export bans imposed by the Government of Vietnam in 2007 and 2008. In addition, the time of the survey, which was in May and September of 2006, is the closest to the time frame of VIRIO 2005. Both income and expenditure information was collected from 9,189 of households, or roughly 0.05 percent of all households in Vietnam. VHLSS 2006 is a multi-stage stratified random sample, split by urban and rural households.

We follow Deaton (1989) in using non-parametric kernel regressions in our analysis of household demand and supply patterns for rice, as well as household welfare impacts as a result of an increase in the world price of rice. This approach places a flexible curve on an (x, y) scatterplot with no parametric restrictions on the form of the curve (Cameron and Trivedi, 2005), thereby providing easily comprehensible descriptions of data across the population.

Throughout this study, we use household expenditure per person as our measure of household living standards. It could be argued that income, rather than expenditure, is a better measure of welfare. However, income can be difficult to measure in a developing country like Vietnam, its large amount of unreported income. By contrast, household expenditures can be measured in an internationally accepted way and can be deflated by region-specific cost of living indices. Another advantage of using expenditure as a measure of welfare is that consumption tends to be smoothed in response to income fluctuations over relatively a long period of time (Deaton, 1997). Based on per capita household expenditure, we define households in the lowest quintile, the highest quintile and the middle quintiles as poor, rich and middle expenditure households.

5.4 Results

In this section, we first present basic results for the household rice demand and supply in Vietnam, followed by model results for both the national and sub-national economies drawn from the bottom-up CGE model. We then analyse the distributional and welfare impacts on households via the micro-simulation.

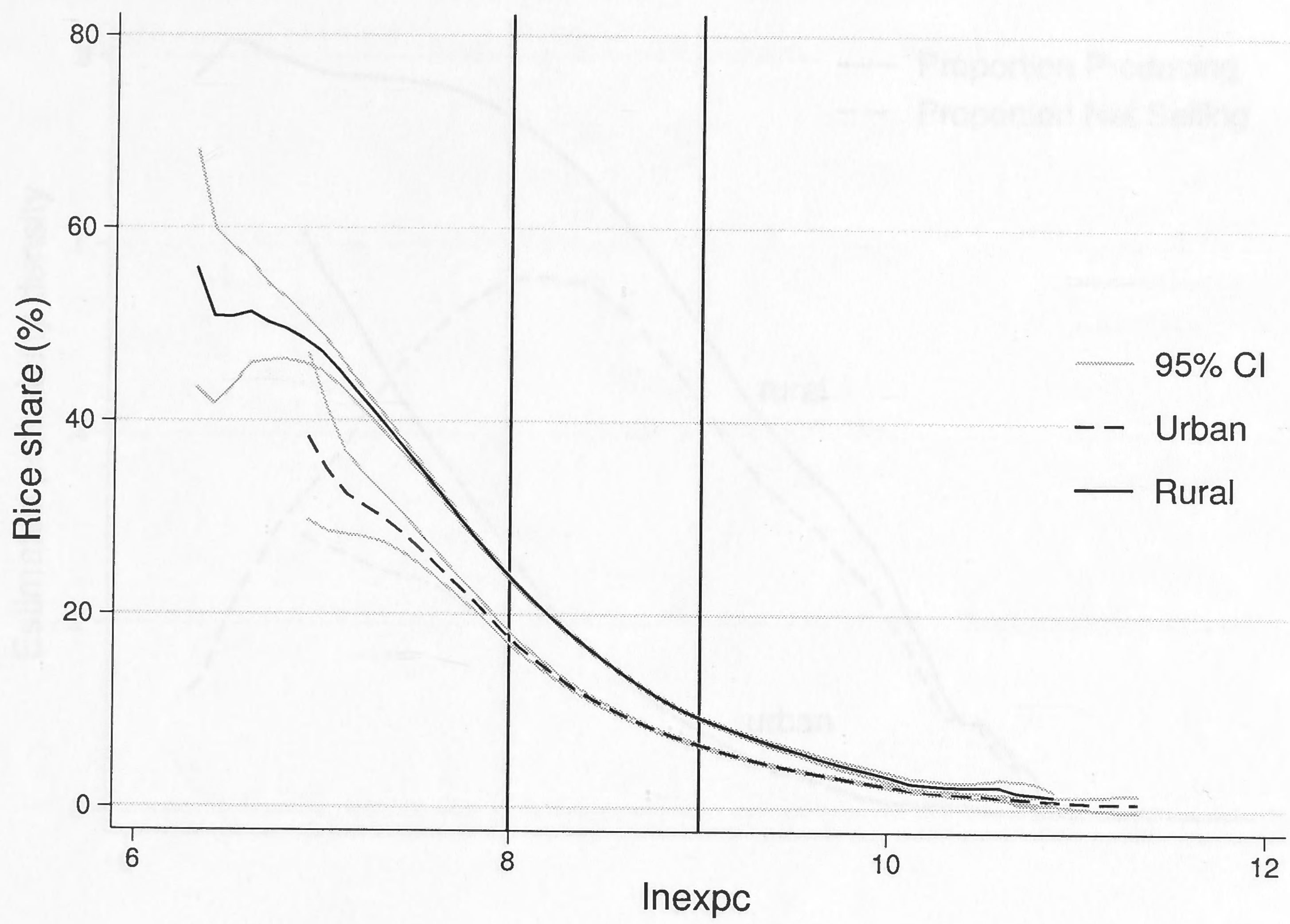
5.4.1 Household rice demand and supply

To analyse household demand and supply patterns for rice, relative to living standards, and how these vary by region, we first calculate rice shares in total expenditure and estimate the probability of being a rice producer and a net rice seller. A rice producer implies that a household produces rice, but may or may not produce enough rice for its own consumption. Later, in section 4.3, we classify households as simply net buyers or sellers of rice, based on the difference between household consumption and production of rice.

Figure 5.2 shows the non-parametric regressions of rice shares on the logarithm of household per capita expenditure, or *lnexpc*. Here, the rice share is the share of rice in household total expenditure. The logarithmic transformation is chosen to reduce skewness of household per capita expenditure (Deaton, 1989). The analysis reveals that rice is an important staple, and the poorer the household the more important is rice in household consumption. For example, poor households spend between 20 to 50 percent of their expenditure on rice, while rich households spend from less than 1 to 10 percent. Rice also accounts for a larger share of rural household expenditure than is the case for urban households, since rural households are generally poorer than their urban counterparts. On these results alone, increases in rice prices will differentially harm poorer households, whether rural or urban, but especially so for rural households. As there is no apparent difference among regions in terms of rice consumption in the data, we do not include an illustration of this result.

Figure 5.3 presents the estimated probability densities of rural and urban households producing rice and those that are net sellers of rice as functions of *lnexpc*. Rice production is clearly an activity of rural households, and especially so for poor households. The solid lines, depicting rural and urban households producing rice, show that nearly 80 percent of poor rural households are rice producers in comparison to 30 to 60 percent of poor urban households. It is also clear that the

FIGURE 5.2: Rice Share Regressions

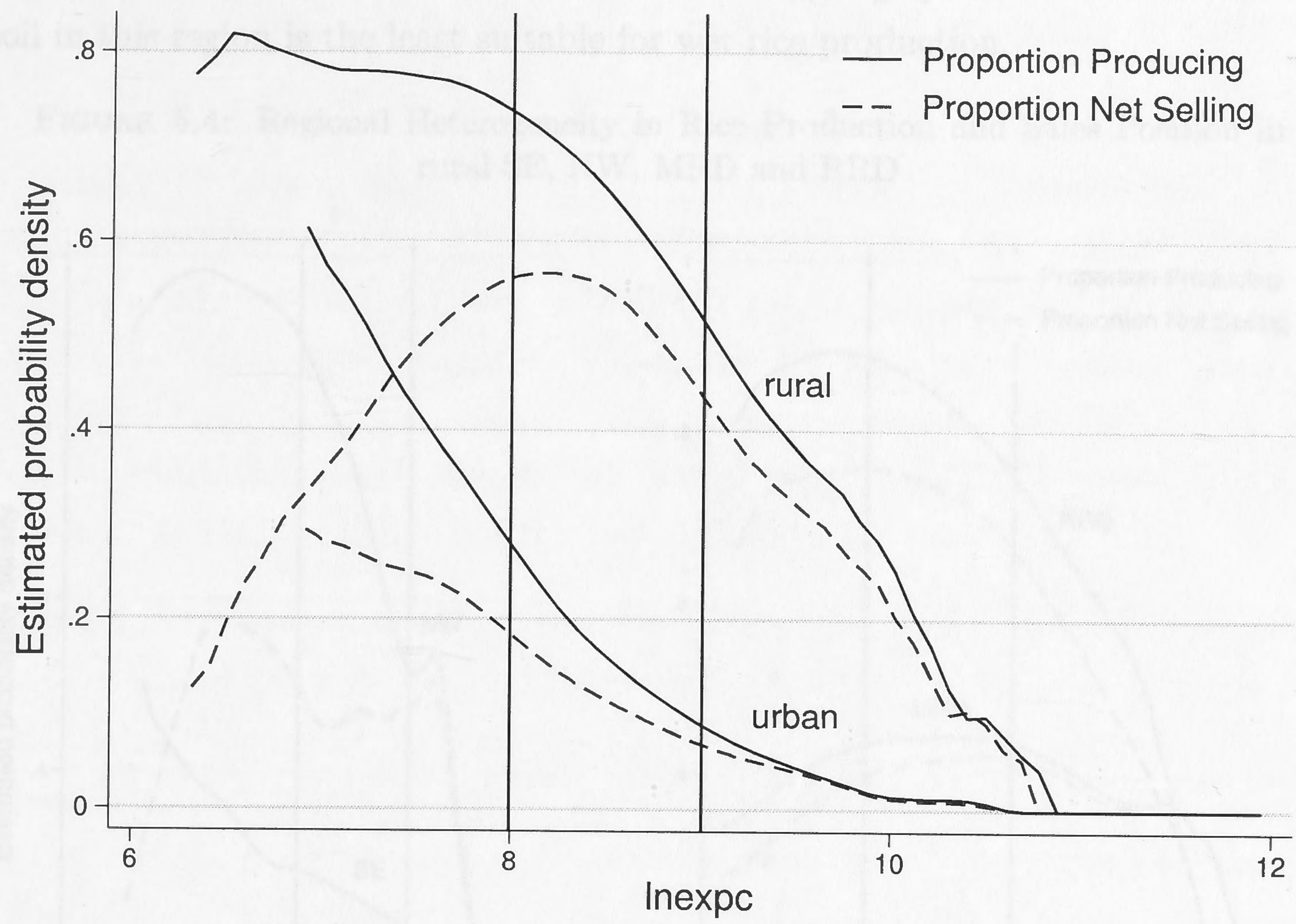


Rice share is the share of rice in a household expenditure. Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel=epanechnikov; degree=0; bandwidth=.2 for urban and 0.16 for rural.

difference between the expected probability of being a net rice seller, the dashed lines, and that of being a rice producer, gets smaller as living standards increase for both rural and urban households. This result suggests a possible lack of suitable land, labour and other necessary materials for poor households to produce rice to meet their own consumption demands. The two solid vertical lines, represent the 20th and 80th percentiles of the distribution.

Figure 5.4 further elaborates the results in Figure 5.3 by regions. Here, we focus only on rural areas where most of rice production occurs, although it is not uncommon in Vietnam for poor urban households to have rice plots in nearby fields. For brevity, we highlight only the most pronounced regional differences (with comparisons for other regions collected in Appendix 5.6.4). The right panel in Figure 5.4, for example, compares the two key rice producing regions, the RRD and the MRD. Households in the RRD are much more likely to produce and sell rice than those in the MRD, and these probabilities fall as their living standards increase. The estimated probabilities of selling and producing rice are relatively stable across the *lnexpc* distribution for MRD households. Furthermore, the gap between those two

FIGURE 5.3: Rice Production and Sales Position

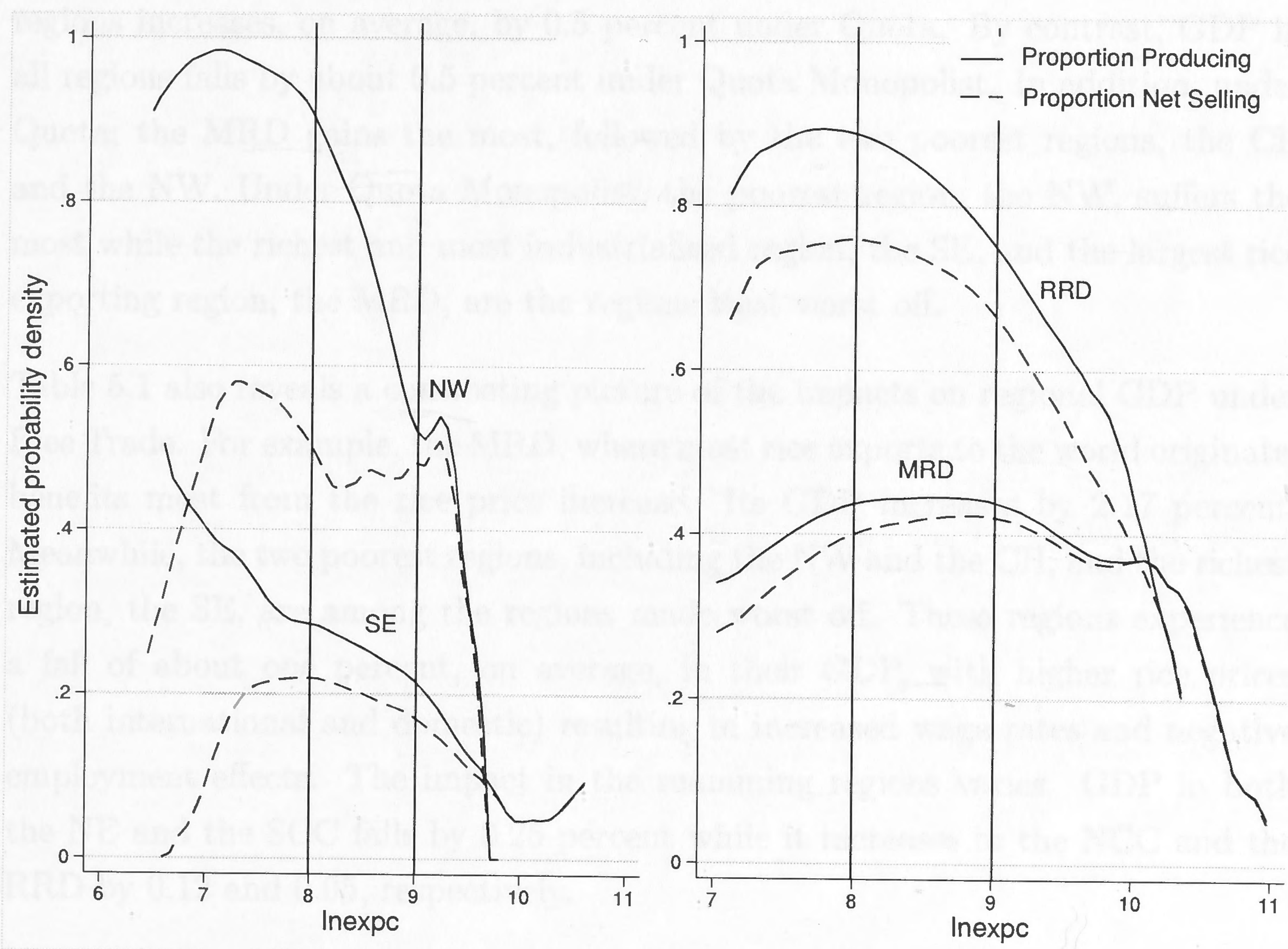


Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel = epanechnikov; degree = 0; banwidth=0.25 and 0.22 for rural proportion producing and selling, respectively; 0.36 for both urban proportion producing and selling.

estimated probabilities is smaller in the MRD. This likely reflects the remnants of land policy in Vietnam over the last three decades: small and non-contiguous plots of land were allocated to households in the rural North, including the RRD, after the dismantling of agricultural collectives, to ensure equity, hindering land consolidation and accumulation and leading to rice production largely on a small scales or even at subsistence levels. Rice farms in MRD, on the other hand, are larger and more consolidated, allowing for mechanisation (Kompas et al., 2012). MRD rice farmers are thus more likely to produce at a larger scale.

The left panel in Figure 5.4 compares Vietnam's richest and most industrialised region, the South East (SE), with the poorest and the most remote region in the North East (NE), near the border with China. There is a sharp contrast between these two regions. Almost all poor households in the NE produce rice, compared to 30 to 50 percent of poor households in the SE. Households in the SE have access to manufacturing jobs in factories and small industries, throughout the region, while their counterparts in the NE have almost no off-farm job opportunities. Although many NE households produce rice, unfortunately, a large number cannot produce enough rice to meet their own household demands, largely due to the fact that the soil in this region is the least suitable for wet rice production.

FIGURE 5.4: Regional Heterogeneity in Rice Production and Sales Position in rural SE, NW, MRD and RRD



Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel = epanechnikov; degree = 0; banwidth=0.31, 0.42, 0.29 and 0.35 for both rural proportion producing and selling in RRD, MRD, NW and SE, respectively.

5.4.2 Results from the bottom-up CGE model

The model results presented in Table 5.1 show the impact of a 30 percent increase in the world price of rice on national and regional GDP in Vietnam under the different trade scenarios. At the national level, the impact is small. In particular, GDP falls slightly under Free Trade, by 0.06 percent, it increases by 0.6 percent under Quota and falls by 0.37 percent under Quota Monopolist. This small aggregate impact is consistent with results from static CGE models in Vietnam.[10] Furthermore, the slight negative impact under Quota Monopolist is consistent

[10] See Abbott et al. (2009) for a comprehensive review of and predictions from CGE models in Vietnam.

with the actual slow down in the GDP growth rate in Vietnam from 8 percent for the 2004-2007 period to 6 percent in 2008, with high food prices often seen as the most contributing factor (Central Institute for Economic Management, 2010).

At the regional level, Table 5.1 reveals a relatively homogeneous picture of the regional GDP impact under Quota and Quota Monopolist. Indeed, GDP in all regions increases, on average, by 0.5 percent under Quota. By contrast, GDP in all regions falls by about 0.5 percent under Quota Monopolist. In addition, under Quota, the MRD gains the most, followed by the two poorest regions, the CH and the NW. Under Quota Monopolist, the poorest region, the NW, suffers the most while the richest and most industrialised region, the SE, and the largest rice exporting region, the MRD, are the regions least worst off.

Table 5.1 also reveals a contrasting picture of the impacts on regional GDP under Free Trade. For example, the MRD, where most rice exports to the world originate, benefits most from the rice price increase. Its GDP increases by 2.17 percent. Meanwhile, the two poorest regions, including the NW and the CH, and the richest region, the SE, are among the regions made worst off. These regions experience a fall of about one percent, on average, in their GDP, with higher rice prices (both international and domestic) resulting in increased wage rates and negative employment effects. The impact in the remaining regions varies. GDP in both the NE and the SCC falls by 0.25 percent while it increases in the NCC and the RRD by 0.12 and 0.05, respectively.

All together, the results in Table 5.1 suggest that all regions, except the MRD, have the highest increase in their GDP under Quota, and that the MRD, not unsurprisingly, gains the most under Free Trade. Most importantly, Table 5.1 indicates that Quota Monopolist is the worst of all policies for the country and all its regions in terms of GDP.

As shown in Table 5.2, the model results indicate that domestic rice prices increase rapidly under Free Trade, moderately under Quota Monopolist, and fall under Quota, all as expected. It is also apparent that the impact on regional domestic rice prices is very similar for the MRD and the SE, on one hand, and all of the remaining regions on the other. For example, in the MRD and the SE, under Free Trade, domestic rice prices increase by more than 30 percent, the simulated increase in world rice prices, while they fall the most sharply (by about 17 percent) under Quota. For other regions, under Free Trade, domestic rice prices increase by about 26 percent and fall by about 5 percent under Quota. These contrasting regional results are largely driven by the weak market integration between rice

TABLE 5.1: Change in GDP (percent)

Whole country/regions	Free Trade	Quota	Quota Monopolist
Vietnam	-0.06	0.6	-0.37
Regions			
Red River Delta	0.05	0.41	-0.62
North East	-0.25	0.26	-0.72
North West	-0.91	0.78	-0.89
North Central Coast	0.12	0.29	-0.59
South Central Coast	-0.25	0.15	-0.56
Central Highland	-1.3	0.82	-0.54
South East	-0.88	0.6	-0.22
Mekong River Delta	2.17	1.07	-0.18

markets in the north and south rice of Vietnam (Baulch et al., 2008), while the MRD and the SE are highly inter-connected, with resulting co-movements in rice prices. Furthermore, the MRD is the largest rice exporter while the SE is the key rice-processing region. Under trade liberalisation, these two regions are the most exposed to changes in the world rice market, and thus experience the bulk of the changes in the world demand for rice. By contrast, they also tend to generate the largest excess supplies of rice when rice export restrictions are in place, with consequent and significant falls in rice prices.

TABLE 5.2: Change in Regional Domestic Rice Prices (percent)

Whole country/regions	Free Trade	Quota	Quota Monopolist
Red River Delta	27.64	-7.32	12.32
North East	23.94	-4.70	13.62
North West	26.10	-4.02	13.68
North Central Coast	26.16	-4.45	13.57
South Central Coast	24.52	-4.14	13.84
Central Highland	26.78	-5.15	12.70
South East	32.02	-18.27	6.49
Mekong River Delta	34.24	-16.8	6.41

Changes in the Consumer Price Index (CPI) are presented in Table 5.3. Their trend is closely linked with domestic rice prices, since rice is a key component in calculated consumption baskets. Nominal regional wages also change accordingly, as reported in Table 5.4, since throughout the country (and especially for unskilled labour) they are largely indexed to the CPI — a result also consistent with the assumption of fixed regional labour supplies in our model. Table 5.5 indicates changes in regional employment. Importantly, unskilled workers in key rice producing regions, most of whom are poor, are likely to have more employment and (using Table 5.4) much higher wages under Free Trade. These workers are more likely be involved in rice production and rice-related industries that expand with Free Trade and other agricultural sectors. On the other hand, higher wages and shifts in labour to key rice growing areas within a region generate falls in employment in other industries and areas. In the other two trade scenarios, more employment, both skilled and unskilled, is created under Quota, compared to Quota Monopolist in all regions.

TABLE 5.3: Change in regional Consumer Price Index (percent)

	Free Trade		Quota		Quota Monopolist	
	Urban	Rural	Urban	Rural	Urban	Rural
Red River Delta	3.14	5.48	-0.87	-1.66	0.92	1.79
North East	3.27	5.39	-0.96	-1.64	1.05	1.95
North West	4.58	7.83	-1.5	-2.71	1.34	2.43
North Central Coast	3.29	5.19	-0.84	-1.43	0.96	1.69
South Central Coast	2.22	3.9	-0.36	-0.66	0.85	1.76
Central Highland	3.61	6.42	-1.4	-2.65	0.77	1.46
South East	2.16	4.44	-0.93	-2.2	0.37	0.83
Mekong River Delta	4.17	6.78	-1.29	-2.58	0.56	1.05

TABLE 5.4: Change in Regional Wages (percent)

Whole country/regions	Free Trade	Quota	Quota Monopolist
Red River Delta	4.93	-1.06	0.99
North East	4.74	-1.26	1.17
North West	6.5	-1.9	1.59
North Central Coast	5.12	-1.09	1.1
South Central Coast	3.17	-0.43	0.95
Central Highland	4.5	-1.6	0.82
South East	2.31	-0.91	0.36
Mekong River Delta	8.37	-1.33	0.75

The overall results can be explained by a combination of features in Vietnam's economic structure and the movements in output and input prices. First, the small impact on national GDP in Vietnam from the world rice price increase is consistent with the small share of the value of rice production in Vietnam's

TABLE 5.5: Change in Regional Employment (percent)

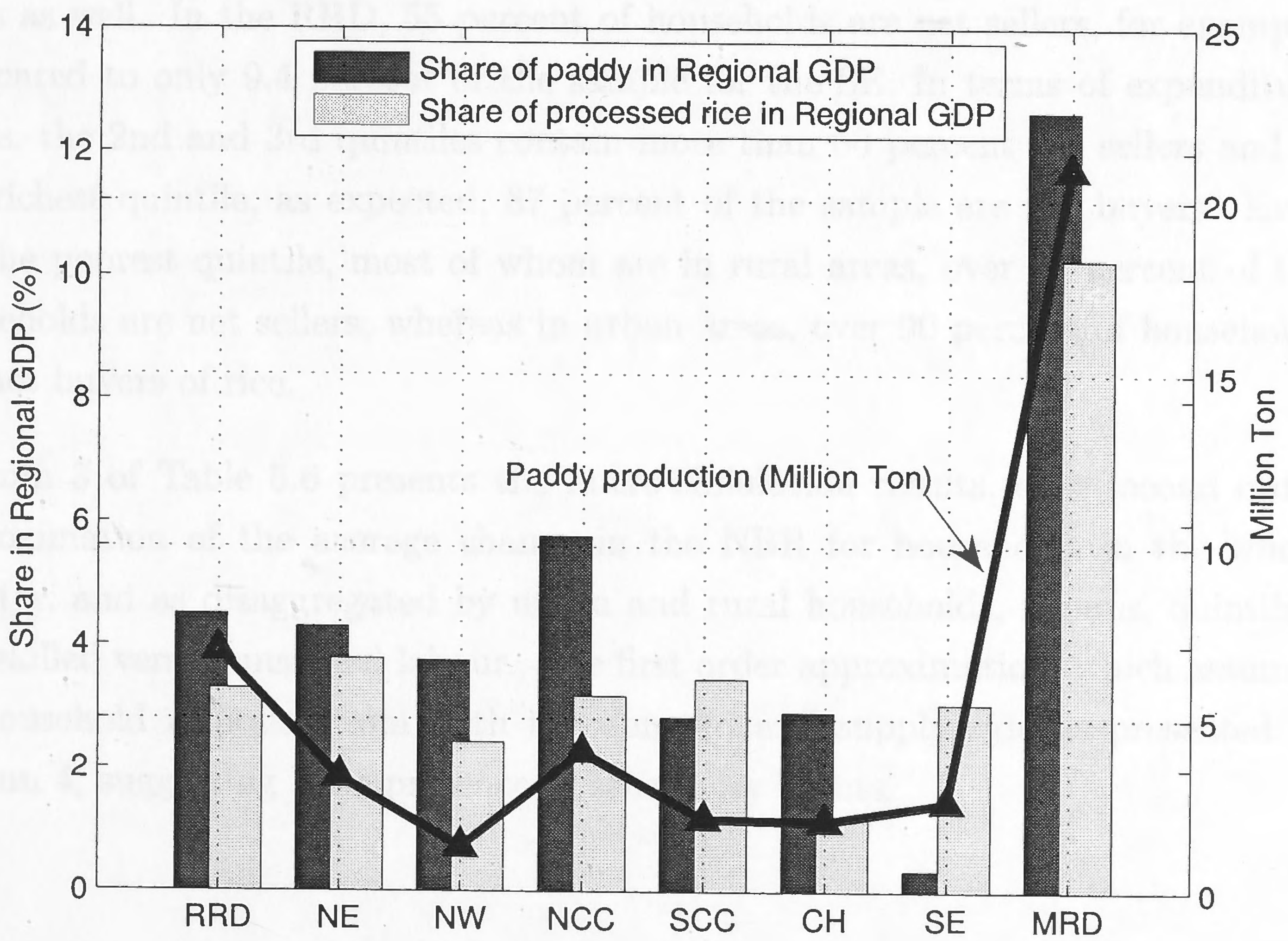
	Free Trade		Quota		Quota Monopolist	
	Skilled	Unskilled	Skilled	Unskilled	Skilled	Unskilled
Red River Delta	-0.75	1.44	0.63	0.58	-0.83	-0.93
North East	-0.92	0.31	0.41	0.34	-0.88	-1.05
North West	-1.91	-0.84	1.06	1.13	-1.06	-1.19
North Central Coast	-0.45	1.06	0.4	0.4	-0.73	-0.86
South Central Coast	-0.69	0.33	0.22	0.19	-0.76	-0.8
Central Highland	-1.8	-1.53	1.04	1.2	-0.67	-0.74
South East	-1.5	-1.31	0.99	1.02	-0.36	-0.37
Mekong River Delta	2.16	5.17	1.56	1.73	-0.34	-0.28

economy. Although distributional and welfare effects can be large, rice production itself accounts for only (roughly) 10-15 percent of GDP (World Bank, 2009c).

Second, a slight fall in GDP under Free Trade is also consistent with the fact that gains in the expanding rice sector are offset by losses in the rest of the economy. In particular, when domestic rice prices increase in all regions under Free Trade (Table 5.3), producers of rice expand production with enhanced profitability. Since they are largely labour-intensive producers, they also generate upward pressure on labour demand and wages, especially for unskilled workers. Furthermore, increases in rice prices also increase the CPI (Table 5.3) and nominal wages (Table 5.4). Labour, an important input, thus becomes more expensive, imposing a loss to sectors where output prices have not risen, or have not risen as much as wages, unless there are substantial opportunities to substitute capital for labour. The net result is a (slight) fall in GDP due to the increased production of low-valued rice compared to other parts of the economy. When export trade is limited, alternatively, as is the case under Quota, domestic rice prices fall, leading to lower nominal wages in all regions (Tables 5.2, 5.3 and 5.4). Resources are therefore shifted away from the rice sector to other sectors, which account for a larger share of national output, and a larger (albeit small) positive impact on national GDP. Quota Monopolist again represents the worst case scenario, since domestic rice prices increase with no compensating gains to the rice exporting regions due to export controls.

Finally, the heterogeneity in the regional impacts, especially under Free Trade, can be explained by three factors. These factors include regional comparative advantage in rice production, the regional contribution of rice production to GDP, and the direction of price and wage movements. As shown in Figure 5.5, rice production is centred largely in the MRD, contributing the highest share to regional GDP. Therefore, the regional impact of a higher rice price is the most pronounced in this region. In contrast, rice production accounts for less than five percent (in almost all cases) of regional GDP in other regions, with much smaller output throughout. In addition, higher prices from Free Trade also generate employment effects in other regions. For example, the SE, the most industrialised region, is also among the most worst off under Free Trade largely because its industrial and non-rice sectors have higher labour costs as a result.[11] The loss for non-rice sectors thus far outweighs the gain in the expanding rice sector in this region. Finally, not all regions in Vietnam produce enough rice for their own demands. In fact, the NW and the CH have to import rice, mainly from the MRD. The NW, the most remote region, also has to pay higher margin costs. Consequently, these two regions are made relatively worse off when domestic rice prices increase.

FIGURE 5.5: Regional GDP Shares of Paddy and Processed Rice and Regional Paddy Production



Regional GDP shares of Paddy and Processed Rice are from VIRO 2005; Regional Paddy Production is from Statistical Yearbook of Vietnam 2009.

[11]See Central Institute for Economic Management (2010, pp. 92-94) on labour movements from industrial regions, especially the SE, to rural areas in 2008.

5.4.3 Distributional and welfare impacts

As mentioned, although the impacts of changes in the world price of rice are small nationally, in terms of GDP, there are significant distributional impacts on regional household welfare. With this in mind, it is important to first note that there are a large number of net sellers of rice in Vietnam, and that this measure varies by region and expenditure levels. Here, households are classified as net rice sellers or buyers based on household differences between rice production and consumption. A household is defined as a net rice seller if the value of its rice production is higher than the value of its rice consumption; and alternatively, as a net rice buyer. Rice sufficient households are those where the value of produced rice exactly equals the value of consumption. As only 13 households in the sample of more than 9000 households are rice sufficient, their results are not presented here for simplicity.

Household net seller or net buyer positions in rice as well as their (mean) percentages in the household data are shown in column 2 of Table 5.6. In terms of all households, 40 percent are net sellers, and in rural areas over 50 percent of households are net sellers. There are also regional and expenditure level variations as well. In the RRD, 55 percent of households are net sellers, for example, compared to only 9.4 percent of the sample for the SE. In terms of expenditure levels, the 2nd and 3rd quintiles contain more than 50 percent net sellers and in the richest quintile, as expected, 87 percent of the sample are net buyers. Even for the poorest quintile, most of whom are in rural areas, over 48 percent of the households are net sellers, whereas in urban areas, over 90 percent of households are net buyers of rice.

Column 3 of Table 5.6 presents the micro-simulation results, as a second order approximation of the average change in the NBR for households in the whole country, and as disaggregated by urban and rural households, regions, quintiles, and skilled versus unskilled labour. The first order approximation, which assumes no household response from both the demand and supply side, is presented in column 4, suggesting little presence of sensitivity effects.

TABLE 5.6: Household net position in rice and net benefit ratio (%)

Categories	% of all hh	% of hh in catergories		Second order approximation NBR			First order approximation NBR		
		Net Seller	Net Buyer	Free Trade	Quota	Quota Monopolist	Free Trade	Quota	Quota Monopolist
All	100.0	40.1	59.7	4.7	1.2	-1.3	4.4	1.1	-1.4
Urban	25.1	8.8	90.8	0.8	0.3	-0.4	0.6	0.3	-0.4
Rural	74.9	50.6	49.3	6.3	1.5	-1.7	5.8	1.5	-1.7
Red River Delta	21.2	55.2	44.8	4.0	0.7	-1.3	3.7	0.7	-1.3
North East	14.3	46.8	53.2	3.3	0.7	-2.2	3.0	0.7	-2.2
North West	4.7	35.8	64.2	1.7	2.5	-2.6	1.1	2.5	-2.7
North Central Coast	11.0	51.0	49.0	5.2	1.0	-2.0	4.8	1.0	-2.1
South Central Coast	9.3	44.3	55.3	2.6	0.2	-1.5	2.4	0.2	-1.5
Central Highlands	6.3	22.1	77.8	-0.3	2.3	-1.5	-0.8	2.3	-1.5
South East	12.9	9.4	89.8	-0.2	1.0	-0.4	-0.4	0.9	-0.4
Mekong River Delta	20.3	35.0	65.0	12.5	2.1	-0.9	12.0	2.1	-0.9
Poorest Quintile	19.0	48.3	51.7	4.8	2.4	-2.8	4.2	2.3	-2.9
2nd Quintile	19.7	55.1	44.8	7.0	1.5	-1.8	6.5	1.5	-1.9
3rd Quintile	20.6	51.1	48.9	6.5	1.3	-1.4	6.1	1.2	-1.4
4th Quintile	20.6	35.7	64.3	4.7	0.8	-0.8	4.4	0.8	-0.8
Richest Quintile	20.1	12.3	87.0	1.3	0.2	-0.3	1.2	0.2	-0.3
Skilled	56.3	34.0	65.7	4.0	1.0	-1.1	3.6	0.9	-1.1
Unskilled	43.8	46.6	53.4	5.8	1.5	-1.6	5.3	1.4	-1.7

The micro-simulation shows, on average, that the NBR of households in the whole country increases by 4.7 percent under Free Trade and 1.2 percent under Quota. It is negative or falls by -1.3 percent under Quota Monopolist. The differences are explained by the fact that rising rice prices both benefit net sellers of rice and, given the results of the CGE modelling, increase nominal wages and incomes.[12] Quota Monopolist, or the status quo, once again, stands out as an inferior policy from the household's perspective, with households in all disaggregated categories invariably worse off. Under Quota, all households gain, with considerable variance in results compared to Free Trade, depending on region in particular.

5.4.3.1 Distributional and welfare impacts: rural versus urban

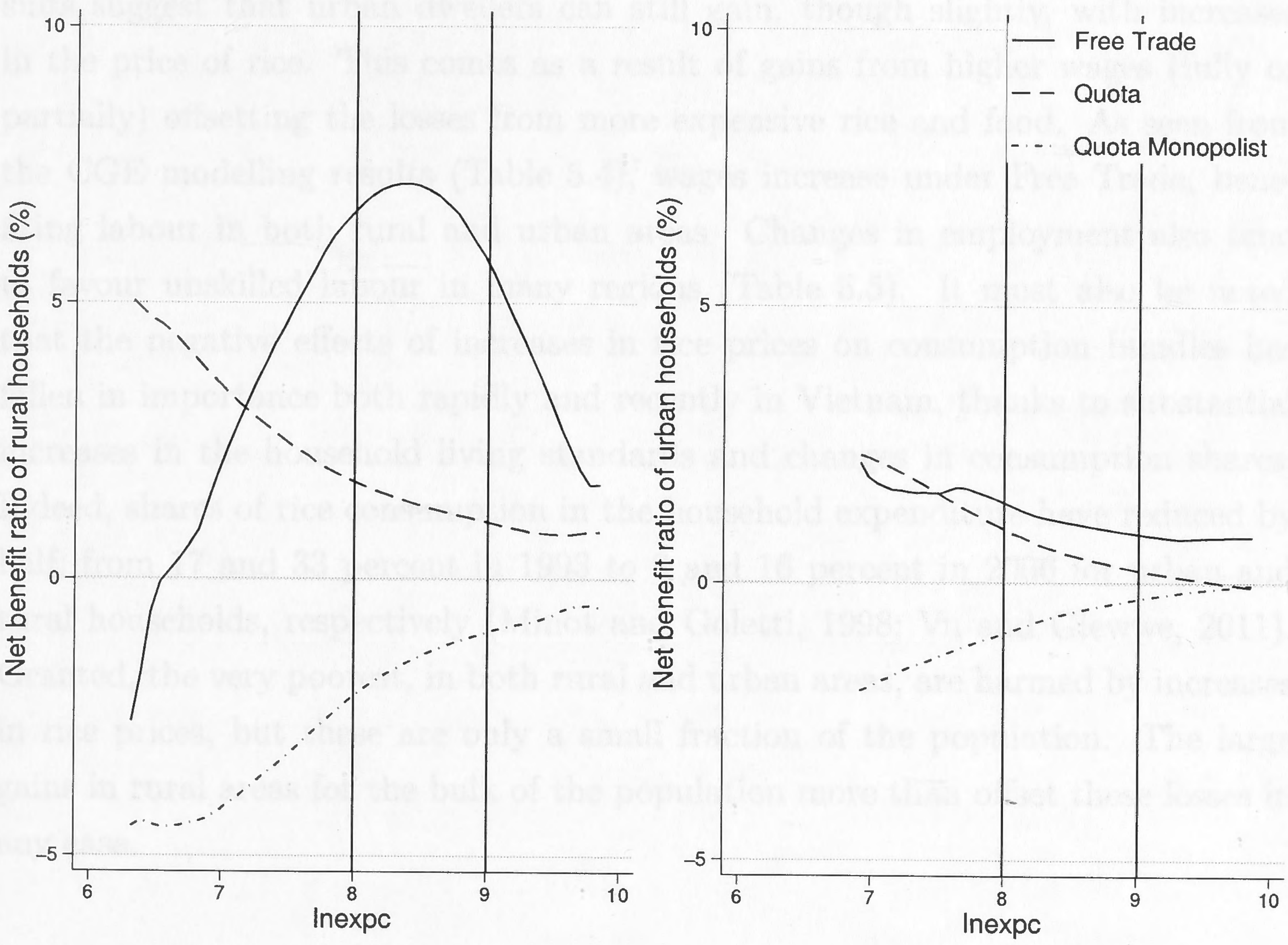
Differences between rural and urban households are profound under different trade scenarios. Figure 5.6 shows changes in the NBR across the *lnexpc* distribution for rural and urban households by scenario. Solid vertical lines are the 20th and 80th percentile of the distribution.[13] Although Free Trade dominates the other two scenarios across most of the distribution, for both rural and urban households, it shows especially enhanced gains in rural areas for the bulk of the population. In particular, in rural areas, the gain under Free Trade is substantially higher than that under Quota Monopolist across the entire *lnexpc* distribution, and also much higher than under Quota for all rural households, except for those in the lowest 2nd percentile. On average, the gain under Free Trade in rural Vietnam is 6.3 percent compared with 1.5 percent and -1.7 percent, under Quota and Quota Monopolist. The gain under Free Trade is highest for the middle-expenditure households. For the very poor rural households, on the other hand, only the 0.12th and lower percentile are worse off under Free Trade. These poorest households often lack

[12] Our results for Vietnam as a whole support aggregative and partial equilibrium findings in (Minot and Goletti, 1998), where it is shown that trade liberalisation increases average real income, as well as findings in Vu and Glewwe (2011) and Phung and Waibel (2010), where it is argued that higher food prices increase average Vietnamese household welfare.

[13] With kernel density estimation there is a genuine trade-off between avoiding bias and reducing the variance of the estimate. In our sample, there is a large variance in the NBR at the top end of the *lnexpc* distribution. This is plausible since very rich households tend to have much more variety in consumption and (to a lesser extent) production, especially among rural households. In any case, we are interested in households where rice production and/or consumption is important. For the richest two percent of households, rice consumption accounts for less than two percent of their household expenditure. These households are also less than two percent likely to be involved in rice production. Hence, to focus our attention and to avoid over-smoothing our estimates for households by increasing the bandwidth in the lower end of *lnexpc*, we exclude the richest two percent households from the sample in our kernel regression estimation of NBR on *lnexpc*.

working-labour, have sick people or many young children, and have no or little fertile land.

FIGURE 5.6: Change in Household Welfare: Rural versus Urban



Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel = epanechnikov, degree = 0; bandwidth=(0.27; 0.29; 0.2) and (0.4; 0.4; 0.37) for rural and urban under Free Trade, Quota and Quota Monopolist, respectively. Two percent richest households are excluded.

Urban households also gain the most under Free Trade, albeit on average only by 0.8 percent. In Figure 5.6, all urban households are shown to do better under Free Trade than under Quota Monopolist and only 4 percent of poorest households are expected to do better under Quota than under Free Trade. With as much as 90 percent of households as net rice buyers, the gain for urban households under Free Trade may come as a surprise. Indeed, our results differ from the literature in this regard, with previous findings on Vietnam and for other countries indicating that urban households as net rice buyers are generally worse off with increases in rice prices (e.g. Minot and Goletti, 1998; Ivanic and Martin, 2008; Vu and Glewwe, 2011; Phung and Waibel, 2010; Warr, 2008). While these results seem sensible, they often ignore the impact of changes in rice prices on wages and employment, due to the use of partial equilibrium modelling (Minot and Goletti, 1998; Ivanic and Martin, 2008; Vu and Glewwe, 2011; Phung and Waibel, 2010), or the lack of detailed household data in a general equilibrium framework (Warr, 2008, 2005).

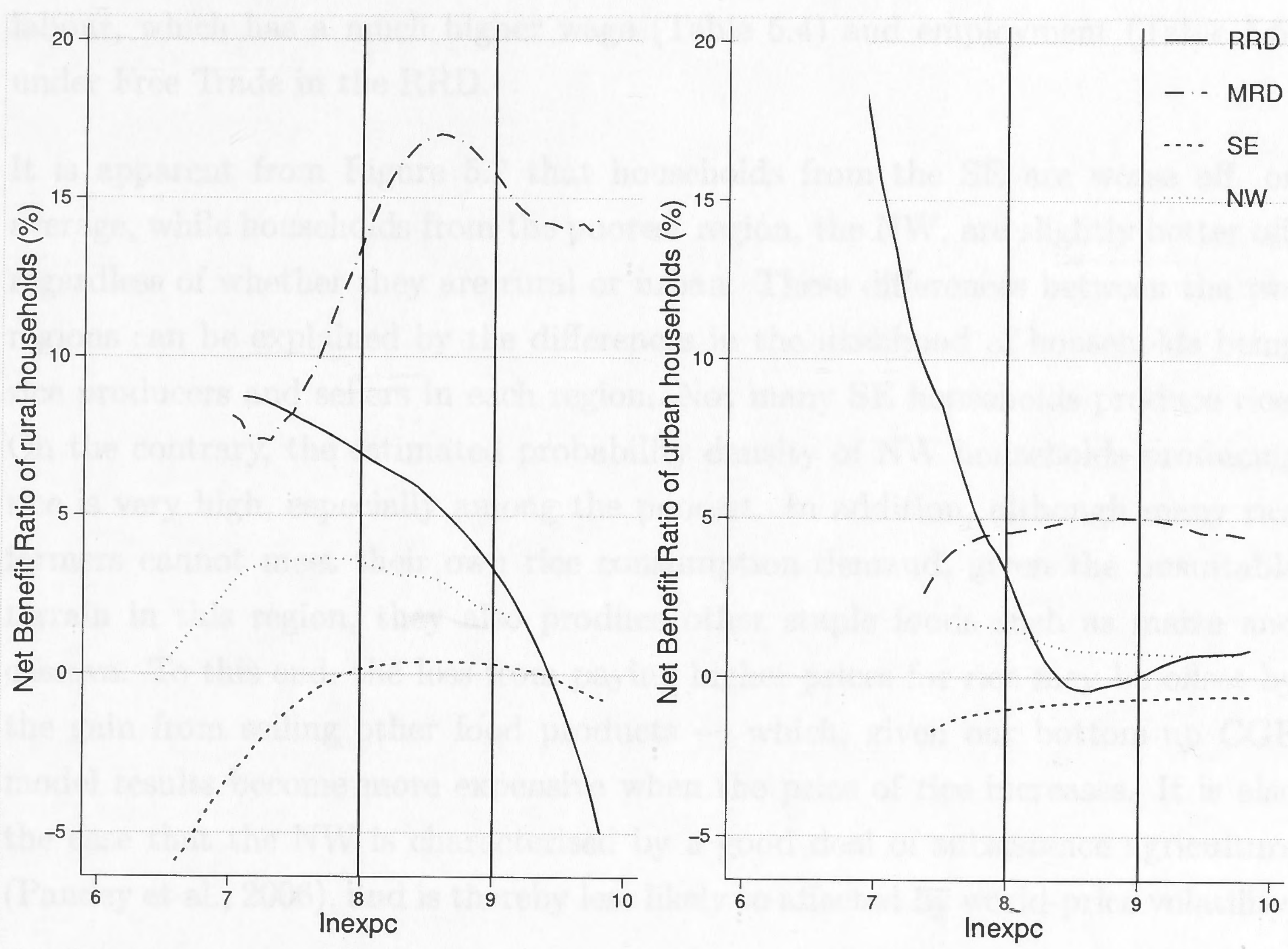
Combining both the bottom-up CGE results and the micro-simulations, our results suggest that urban dwellers can still gain, though slightly, with increases in the price of rice. This comes as a result of gains from higher wages (fully or partially) offsetting the losses from more expensive rice and food. As seen from the CGE modelling results (Table 5.4), wages increase under Free Trade, benefiting labour in both rural and urban areas. Changes in employment also tend to favour unskilled labour in many regions (Table 5.5). It must also be noted that the negative effects of increases in rice prices on consumption bundles has fallen in importance both rapidly and recently in Vietnam, thanks to substantial increases in the household living standards and changes in consumption shares. Indeed, shares of rice consumption in the household expenditure have reduced by half, from 17 and 33 percent in 1993 to 8 and 16 percent in 2006 for urban and rural households, respectively (Minot and Goletti, 1998; Vu and Glewwe, 2011). Granted, the very poorest, in both rural and urban areas, are harmed by increases in rice prices, but these are only a small fraction of the population. The large gains in rural areas for the bulk of the population more than offset these losses in any case.

5.4.3.2 Distributional and welfare impacts: regional differences

In this section, we focus on distributional impact of Free Trade only, since Free Trade is (for the most part) the preferred policy scenario. Figure 5.7 compares the distributional impacts in the key rice producing regions, the RRD and the MRD, the richest industrial region, the SE, and the poorest region, the NW. Not surprisingly, the gain from export liberalisation is largest for the key rice exporting regions. On average, MRD households are expected to gain 12.5 percent. The gain for rural households in the MRD is especially impressive, ranging from 7 to 16 percent, while their urban counterparts gain about 5 percent across the *lnexpc* distribution. While the gain for rural households is relatively obvious given MRD's position in rice production and export trade, the gain for urban dwellers in MRD is attributed to these households being involved in rice trading and transport activities, as well as the general increase in wages in this region.

The gain in the second largest rice producing region, the RRD, alternatively, exhibits a pro-poor pattern with the gain diminishing with living standard, especially

FIGURE 5.7: Change in Household Welfare: RRD, MRD, SE and NW under Free Trade



Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel = epanechnikov, degree = 0, bandwidth=(0.43; 0.27; 0.46; 0.3) and (0.27; 0.47; 0.56; 0.49) for rural and urban in RRD, MRD, SE and NW, respectively. Two percent richest households are excluded.

for rural RRD households. This pattern is driven by the fact that the probability of being a rice farmer and a net rice seller falls as the living standard increases in rural RRD. Furthermore, given the low productivity of rice production in the RRD, compared to the MRD — due largely to fragmentation of rice farms or to the presence of small and non-contiguous plots (Kompas et al., 2012) — better-off rural households in this region tend to diversify their incomes away from rice production and agriculture in general.

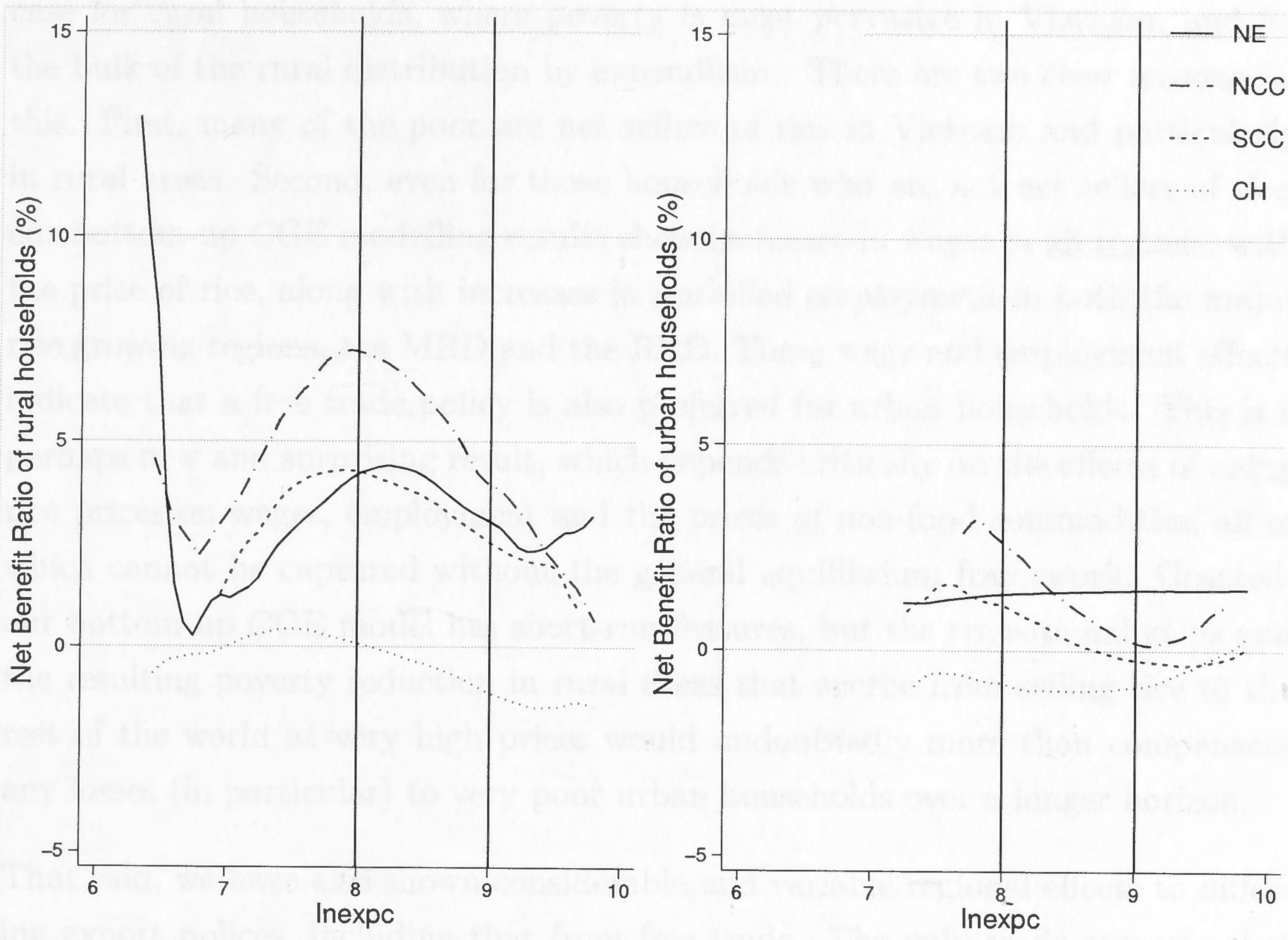
For urban RRD households, the gain from rising rice prices is particularly high for poor households and flattens-out among the middle and rich quintiles. There are two likely explanations for this phenomenon. First, although urbanisation is rapid throughout Vietnam, in this region a large number of administratively classified 'urban' households are actually residing in semi-rural areas, with many households working small rice fields near the edges of urban centres. These households are more likely to be involved in rice production and thus benefit from higher prices. It is also not unusual for urban households to have plots of rice land nearby. Second, the wage effect again matters. Poor urban households normally provide unskilled labour, which has a much higher wage (Table 5.4) and employment (Table 5.5) under Free Trade in the RRD.

It is apparent from Figure 5.7 that households from the SE are worse off, on average, while households from the poorest region, the NW, are slightly better off, regardless of whether they are rural or urban. These differences between the two regions can be explained by the differences in the likelihood of households being rice producers and sellers in each region. Not many SE households produce rice. On the contrary, the estimated probability density of NW households producing rice is very high, especially among the poorest. In addition, although many rice farmers cannot meet their own rice consumption demand, given the unsuitable terrain in this region, they also produce other staple foods such as maize and cassava. To this end, the loss from paying higher prices for rice may be offset by the gain from selling other food products — which, given our bottom-up CGE model results become more expensive when the price of rice increases. It is also the case that the NW is characterised by a good deal of subsistence agriculture (Pandey et al., 2006), and is thereby less likely to affected by world-price volatility.

Figure 5.8 compares the second poorest region, the CH, with the NE, the NCC and the SCC. Households in the CH are clearly worse off, in comparison. This result is plausible given that the CH can only meet as much as 30 percent of its rice demand from local production, with its soil much more suitable for industrial crops such as coffee. On the other hand, the gain from trade liberalisation is prevalent throughout rural areas of all other regions, and highest in the NCC given its net seller position in rice production. The NE, the NCC and the SCC exhibit a similar pattern, with a gain from Free Trade that accrues mostly to rural households, reaching its peak around the 20th percentile in the *lnexpc* distribution.

The gain for urban households in the NCC and the SCC also displays a pro-poor pattern, while it appears more uniformly spread in the NE. The similarity between the NCC and the SCC may be explained by similar economic and development characteristics between the two regions. Both are emerging regions with sea ports, supporting a number of international trading and service activities in urban areas. Urban NE, on the other hand, is relatively less developed, and thus the disparity in expenditures in urban areas is far less pronounced.

FIGURE 5.8: Change in Household Welfare: NE, NCC, SCC and CH under Free Trade



Two solid vertical lines are at the 20th and 80th percentiles of the *lnexpc* distribution. Kernel = epanechnikov, degree = 0, bandwidth=(0.19; 0.29; 0.32; 0.45) and (0.63; 0.32; 0.34; 0.42) for rural and urban in NE, NCC, SCC and CH, respectively. Two percent richest households are excluded.

5.5 Concluding Remarks

This paper analyses Vietnam's rice export policy in the context of rising world rice prices. Bringing together insights from a bottom-up CGE model and a micro-simulation on household data, we provide a clear analysis of Vietnam rice export policy at the aggregate and household level. Perhaps most importantly, we show that one of the key arguments for rice export controls or export bans in the face of rising prices — the desire to protect the poor from welfare losses due to higher rice prices — does not hold in Vietnam.

On the surface, the poor appear to be highly vulnerable to rice price increases. The share of rice in consumption bundles is higher the poorer the household, and especially so in rural areas where most of the poor reside. Urban households are also not likely to be rice producers and thus cannot directly benefit from higher prices for rice.

Nevertheless, our results show that a free trade export policy largely benefits the poor in Vietnam, for both rural and urban households. This is especially the case for rural households, where poverty is most pervasive in Vietnam, and for the bulk of the rural distribution by expenditure. There are two clear reasons for this. First, many of the poor are net sellers of rice in Vietnam and particularly in rural areas. Second, even for those households who are not net sellers of rice, our bottom-up CGE modelling results show increases in wages in all regions, with the price of rice, along with increases in unskilled employment in both the major rice growing regions, the MRD and the RRD. These wage and employment effects indicate that a free trade policy is also preferred for urban households. This is a perhaps new and surprising result, which depends critically on the effects of rising rice prices on wages, employment and the prices of non-food commodities, all of which cannot be captured without the general equilibrium framework. Granted, our bottom-up CGE model has short-run features, but the transitional gains and the resulting poverty reduction in rural areas that accrue from selling rice to the rest of the world at very high prices would undoubtedly more than compensate any losses (in particular) to very poor urban households over a longer horizon.

That said, we have also shown considerable and variable regional effects to differing export polices, including that from free trade. The only trade scenario that generates uniformly worst-case outcomes is the policy that mimics the status quo in Vietnam, one with export controls and regionally fragmented and monopoly-like influence over rice prices and the rice market. A third policy scenario, a quota policy designed to keep domestic rice prices low, is preferable to free trade but only for the lowest 2nd percentile of the rural expenditure distribution. For those very poor, who usually do not grow rice or provide substantial labour services, direct support from the government using a free trade policy, such as that provided by the 'Hunger Eradication and Poverty Reduction Program' in Vietnam, is essential.

5.6 Appendices

5.6.1 Regions and provinces in Vietnam

Region	Provinces
Red River Delta	Ha Noi, Hai Phong, Vinh Phuc, Ha Tay, Bac Ninh, Hai Duong, Hung Yen, Ha Nam, Nam Dinh, Thai Binh, Ninh Binh
North East	Ha Giang, Cao Bang, Lao Cai, Bac Can, Lang Son, Tuyen Quang, Yen Bai, Thai Nguyen, Phu Tho, Bac Giang, Quang Ninh
North West	Lai Chau, Dien Bien, Son La, Hoa Binh
North Central Coast	Thanh Hoa, Nghe An, Ha Tinh, Quang Binh, Quang Tri, Thua Thien Hue
South Central Coast	Da Nang, Quang Nam, Quang Ngai, Binh Dinh, Phu Yen, Khanh Hoa
Central Highlands	Kon Tum, Gia Lai, Dac Lak, Dac Nong, Lam Dong
South East	Ho Chi Minh City, Ninh Thuan, Binh Phuoc, Tay Ninh, Binh Duong, Dong Nai, Binh Thuan, Ba Ria - Vung Tau
Mekong River Delta	Long An, Dong Thap, An Giang, Tien Giang, Vinh Long, Ben Tre, Kien Giang, Can Tho, Hau Giang, Tra Vinh, Soc Trang, Bac Lieu, Ca Mau

5.6.2 Own price elasticities of demand

Code	Industry	Demand Elasticity	Sources
1	Paddy[a]		
2	Other crops[b]	-0.98	Nguyen et al. (2009)
3	Livestock & Poultry[c]	-1.19	Nguyen et al. (2009)
4	Forestry[a]		
5	Fish Farming[a]		
6	Fishery	-1.40	Dey (2000)
7	Oil & gas	-0.08	Deaton and Muellbauer (1980a)
8	Mining[a]		
9	Processed seafood	-1.05	Nguyen et al. (2009)
10	Processed Rice	-0.41	Nguyen et al. (2009)
11	Other Agricultural Processing[d]	-0.42	UNCTAD (2007)
12	Textiles	-0.69	Han and Wahl (1998)
13	Paper	-0.47	Deaton and Muellbauer (1980a)
14	Wood	-0.35	Adams and Haynes (1980)
15	Rubber	-0.25	UNCTAD (2007)
16	Non-Metallic Mineral Products	-0.47	Deaton and Muellbauer (1980a)
17	Transport Equipment	-1.17	Deaton and Muellbauer (1980a)
18	Metal Products[a]		
19	Other Manufacturing	-0.47	Deaton and Muellbauer (1980a)
20	Electricity & Water	-0.53	Cheesman et al. (2008)
21	Construction	-0.21	Han and Wahl (1998)
22	Transport (Margin)	-1.15	Falvey and Gemmell (1996)
23	Communication	-1.66	Falvey and Gemmell (1996)
24	Trade (Margin)[a]		
25	Financial services	-0.74	Deaton and Muellbauer (1980a)
26	Public Administration	-1.20	Falvey and Gemmell (1996)
27	Hotels & Restaurants	-1.06	Falvey and Gemmell (1996)
28	Other Services	-0.74	Deaton and Muellbauer (1980a)

[a] No consumption

[b] Average of demand elasticity of vegetables, maize, roots, proc.starch, tofu and fruits

[c] Average of demand elasticity of pork, beef, chicken, other poultry, other meat and egg

[d] Average of demand elasticity of milk, butter, cheese, refined sugar, processed coffee, coca, tobacco, vegetable oil and tea

5.6.3 Own price elasticities of supply

Code	Industry	Supply Elasticity	Sources
1	Paddy[a]	0.057	Khiem and Pingali (1995)
2	Other crops[b]	0.019	Khiem and Pingali (1995)
3	Livestock & Poultry[c]	0.205	UNCTAD (2007)
4	Forestry	0.89	Adams and Haynes (1980)
5	Fish Farming	0.76	Campbell (1998)
6	Fishery	0.76	Campbell (1998)
7	Oil & gas[d]		
8	Mining[e]	0.10	Evers et al. (2008)
9	Processed seafood[d]		
10	Processed Rice[a]	0.057	Khiem and Pingali (1995)
11	Other Agricultural Processing[f]	0.25	UNCTAD (2007)
12	Textiles	2.50	Lim (2006)
13	Paper	5.59	Shea (1993)
14	Wood	1.06	Shea (1993)
15	Rubber	0.25	UNCTAD (2007)
16	Non-Metallic Mineral Products	5.49	Shea (1993)
17	Transport Equipment[d]		
18	Metal Products	5.49	Shea (1993)
19	Other Manufacturing	5.49	Shea (1993)
20	Electricity & Water		
21	Construction	1.20	DiPasquale (1999)
22	Transport (Margin)[e]	0.10	Evers et al. (2008)
23	Communication[d]		
24	Trade (Margin)[e]	0.10	Evers et al. (2008)
25	Financial services	0.516	Hancock (1985)
26	Public Administration[d]		
27	Hotels & Restaurants	1.976	Fujii et al. (1985)
28	Other Services [e]	0.10	Evers et al. (2008)

[a] Average of supply elasticities for rice in Vietnams South and North

[b] Average of supply elasticities for maize, cassava and sweet potato, for Vietnams South and North

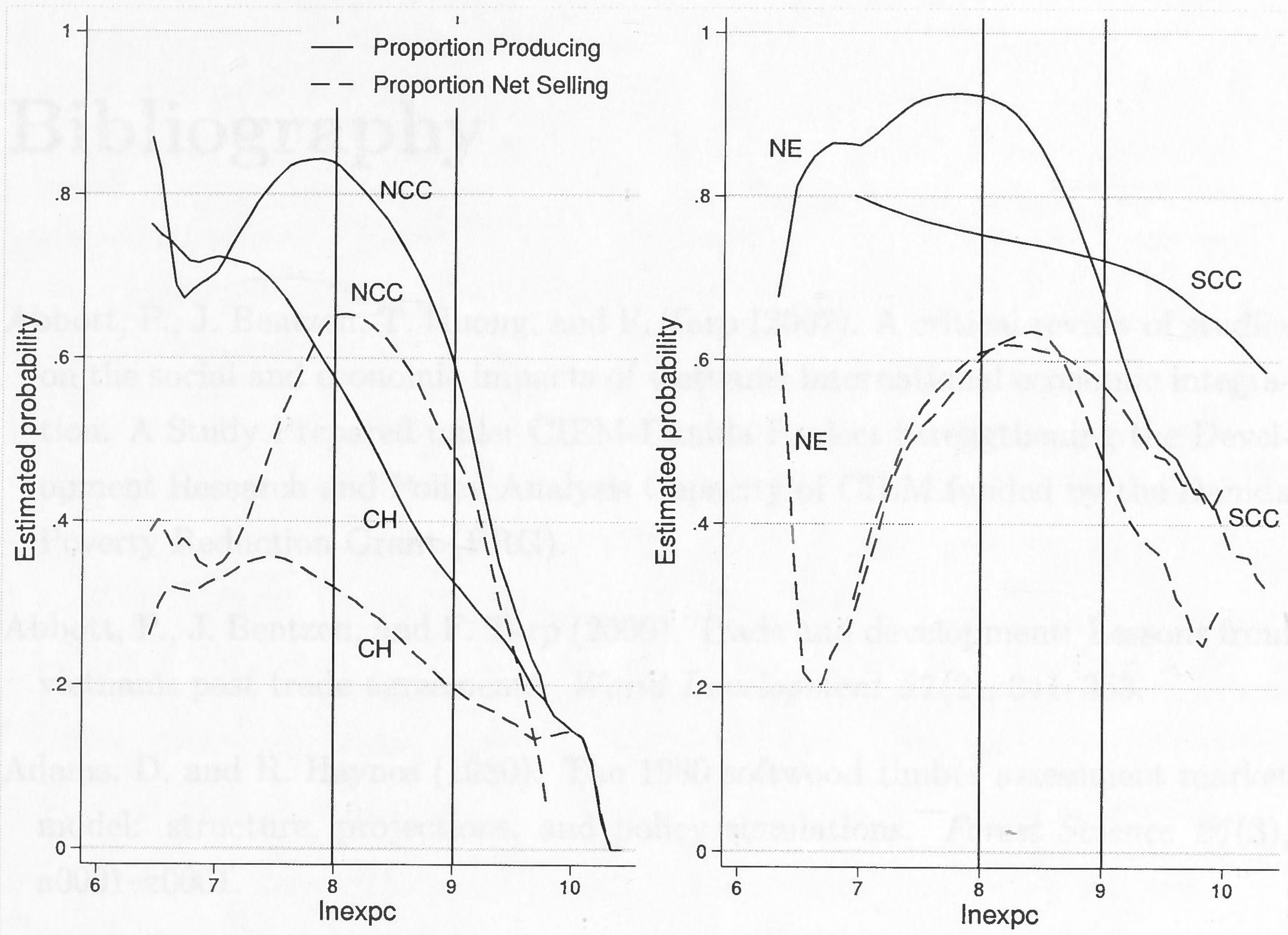
[c] Average of supply elasticities for livestock, bovine meat, pig meat and poultry

[d] No income

[e] Supply elasticity of labour

[f] Average of supply elasticities for milk, butter, cheese, refined sugar, processed coffee, coca, tobacco, vegetable oil and tea

5.6.4 Regional Heterogeneity in Rice Production and Sales Position in rural NCC, CH, SCC and NE



Two solid vertical lines are at the 20th and 80th percentiles of the *lnpcexp* distribution. Kernel = epanechnikov; degree = 0; banwidth=(0.45, 0.4); (0.24, 0.25); (0.24, 0.24) and (0.77, 0.35) for proportion producing and selling in CH, NCC, NE, and SCC, respectively.

Bibliography

Abbott, P., J. Bentzen, T. Huong, and F. Tarp (2007). A critical review of studies on the social and economic impacts of vietnams international economic integration. A Study Prepared under CIEM-Danida Project Strengthening the Development Research and Policy Analysis Capacity of CIEM funded by the Danida Poverty Reduction Grant (PRG).

Abbott, P., J. Bentzen, and F. Tarp (2009). Trade and development: Lessons from vietnams past trade agreements. *World Development 37*(2), 341–353.

Adams, D. and R. Haynes (1980). The 1980 softwood timber assessment market model: structure, projections, and policy simulations. *Forest Science 26*(3), a0001–z0001.

Aigner, D., C. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37.

Audretsch, D. and M. Feldman (1996). R& D spillovers and the geography of innovation and production. *The American Economic Review 86*(3), 630–640.

Baltagi, B. (2005). *Econometric Analysis of Panel Data*. New York: Wiley.

Baltagi, B. H., G. Bresson, and A. Pirotte (2003). Fixed Effects, Random Effects or Hausman–Taylor? A Pretest Estimator. *Economics Letters 79*(3), 361–369.

Bao Duong, P. and Y. Izumida (2002). Rural development finance in vietnam: a microeconometric analysis of household surveys. *World Development 30*(2), 319–335.

Barrett, C. and P. Dorosh (1996). Farmers' welfare and changing food prices: nonparametric evidence from rice in madagascar. *American Journal of Agricultural Economics 78*(3), 656–669.

Battese, G. and T. Coelli (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics 38*(3), 387–399.

Battese, G. and T. Coelli (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics 20*(2), 325–332.

Battese, G. and G. Corra (1977). Estimation of a production frontier model: with application to the pastoral zone of eastern australia. *Australian Journal of Agricultural Economics 21*(3), 169–179.

Battese, G. E. and T. J. Coelli (1993). A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects. Working Papers in Econometrics and Applied Statistics, Department of Econometrics, University of New England, Armidale.

Baulch, B., T. T. K. Chuyen, D. Haughton, and J. Haughton (2007). Ethnic minority development in vietnam. *The Journal of Development Studies 43*(7), 1151–1176.

Baulch, B., H. Hansen, L. Trung, and T. Tam (2008). The spatial integration of paddy markets in vietnam. *Journal of Agricultural Economics 59*(2), 271–295.

Baulch, B., H. T. M. Nguyen, P. T. T. Nguyen, and H. T. Pham (2010). Ethnic Minority Poverty in Vietnam. UK Chronic Poverty Research Centre Working Paper No 169.

Baulch, B., H. T. Pham, and B. Reilly (2012). Decomposing the ethnic gap in rural vietnam, 1993–2004. *Oxford Development Studies 40*(1), 87–117.

Belke, A. and J. Spies (2008). Enlarging the EMU to the east: What effects on trade? *Empirica 35*(4), 369–89.

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human resources 8*, 436–455.

Blomqvist, A. (1986). The village and beyond: Market and public policy (1). In J. F. Ottawa (Ed.), *Scarp Marden Evaluation Baseline Study.* Canada: Canadian International Development.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–50.

Breusch, T., M. Ward, H. T. M. Nguyen, and T. Kompas (2011). On the fixed-effects vector decomposition. *Political Analysis 19*(2), 123–134.

Breusch, T. S., G. E. Mizon, and P. Schmidt (1989). Efficient estimation using panel data. *Econometrica 57*(3), 695–700.

Budd, J. (1993). Changing food prices and rural welfare: A nonparametric examination of the cote d'ivoire. *Economic Development and Cultural Change 41*(3), 587–603.

Bui, T. (1999). *54 Dan Toc o Vietnam va Cac Ten Goi Khac (54 Ethnic Groups in Vietnam and Other Names).* Thanh Nien.

Cameron, A. and P. Trivedi (2005). *Microeconometrics: methods and applications.* Cambridge University Press.

Campbell, H. (1998). How many purse seiners should exploit the western pacific tuna fishery? In U. Chakravorty and J. Sibert (Eds.), *Ocean-Scale Management of Pelagic Fisheries: Economic and Regulatory Issues*, pp. 29–40.

Cantho University (1995). Survey of rice producers 1990–1995. Unpublished data.

Caporale, G. M., C. Rault, R. Sova, and A. Sova (2009). On the bilateral trade effects of free trade agreements between the EU-15 and the CEEC-4 countries. *Review of World Economics 145*(2), 189–206.

Central Institute for Economic Management (2010). *Impacts of International Economic Integration on Vietnamese Economy Three Years after Joining WTO.* Central Institute for Economic Management.

Che, T. N., T. Kompas, and N. Vousden (2001). Incentives and static and dynamic gains from market reform: rice production in vietnam. *Australian Journal of Agricultural and Resource Economics 45*(4), 547–572.

Che, T. N., T. Kompas, and N. Vousden (2006). Market reform, incentives and economic development in vietnamese rice production. *Comparative Economic Studies 48*(2), 277–301.

Cheesman, J., J. Bennett, and V. H. S. Tran (2008). Estimating household water demand using revealed and contingent behaviors: Evidence from vietnam. *Water resources research 44*(11).

Childs, N. and J. Kiawu (2009). Factors behind the rise in global rice prices in 2008. US Department of Agriculture, Economic Research Service.

Chiswick, B. (1991). Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics 13*(2), 149–170.

Chiswick, B. and P. Miller (1995). The endogeneity between language and earnings: International analyses. *Journal of labor economics 9*(2), 246–288.

Chiswick, B., H. Patrinos, and M. Hurst (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change 48*(2), 349–367.

Chu, L. V., T. N. Nguyen, P. H. Phung, T. Q. Tran, and X. T. Dang (1992). *Agricultural Cooperatives in Vietnam: History, Problems, and Prospective.* Hanoi, Vietnam: Su That Publishing House.

Chu, Q. (2008). Rice Production and Food Security in Vietnam. Paper presented at the 33rd Conference of the ASEAN Inter-Association of Science and Economics, November 2008, Hanoi.

Coelli, T. J., D. S. P. Rao, and G. E. Battese (1998). *An Introduction to Efficiency and Productivity Analysis.* Boston: Kluwer.

Committee for Ethnic Minorities (2005). Socio-Economic Development Programme for Extremely Difficult Communes in Ethnic Minority and Mountainous Areas in the Period 2006-2010 (Phase II of the Programme 135). Accessed from http://cema.gov.vn.

Congress of Vietnam (1993). *Land Law.* Hanoi.

Croser, J., P. Lloyd, and K. Anderson (2010). How do agricultural policy restrictions on global trade and welfare differ across commodities? *American Journal of Agricultural Economics 92*(3), 698–712.

Dang, K. S. and C. T. Tran (2008). Role of state-owned enterprises in vietnam's rice markets. In S. Rashid, A. Gulati, and R. Cummings Jr (Eds.), *From parastatals to private trade: lessons from Asian agriculture.* International Food Policy Research Institute (IFPRI).

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford University Press.

Deaton, A. (1989). Rice prices and income distribution in thailand: a non-parametric analysis. *The Economic Journal 99*(395), 1–37.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy.* Johns Hopkins University Press.

Deaton, A. and J. Muellbauer (1980a). An almost ideal demand system. *The American Economic Review 70*(3), 312–326.

Deaton, A. and J. Muellbauer (1980b). *Economics and consumer behavior.* Cambridge University Press.

Deininger, K. and S. Jin (2008). Land sales and rental markets in transition: Evidence from rural vietnam. *Oxford Bulletin of Economics and Statistics 70*(1), 67–101.

Dey, M. (2000). Analysis of demand for fish in bangladesh. *Aquaculture Economics & Management 4*(1-2), 63–81.

DiPasquale, D. (1999). Why don't we know more about housing supply? *The Journal of Real Estate Finance and Economics 18*(1), 9–23.

Dixon, P. and M. Rimmer (2005). Mini-usage: Reducing barriers to entry in dynamic cge modelling.. 8th Annual Conference on Global Economic Analysis, Lübeck, Germany, and June.

DPM (1995). Research project of market and price for rice production in the market economy of vietnam (part one). Hanoi: Unpublished manuscript.

DPM (2002). Research project of market and price for rice production in the market economy of vietnam. Hanoi: Unpublished manuscript.

DPM (2005). Research project of market and price for rice production in the market economy of vietnam. Hanoi: Unpublished manuscript.

Dustmann, C. and F. Fabbri (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal 113*(489), 695–717.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Edmonds, E. and N. Pavcnik (2005). The effect of trade liberalization on child labor. *Journal of International Economics 65*(2), 401–419.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*(397), 171–85.

Evers, M., R. De Mooij, and D. Van Vuuren (2008). The wage elasticity of labour supply: a synthesis of empirical estimates. *De Economist 156*(1), 25–43.

Falvey, R. and N. Gemmell (1996). Are services income-elastic? some new evidence. *Review of Income and Wealth 42*(3), 257–269.

FAO (2007). Yearbook: Fertilizer 1999–2007. Rome.

Feldstein, M. (1974). Errors in variables: A consistent estimator with smaller MSE in finite samples. *Journal of the American Statistical Association 69*(348), 990–96.

Fforde, A. (1996). The institutions of transition from central planning: The case of vietnam. In C. Barlow (Ed.), *Institutions and Economic Change in Southeast Asia*. Cheltenham, U.K.: Edward Elgar.

Forsund, F., C. Lovell, and P. Schmidt (1980). A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics 13*(1), 5–25.

Friedman, J. and J. Levinsohn (2002). The distributional impacts of indonesia's financial crisis on household welfare: A rapid response methodology. *The World Bank Economic Review 16*(3), 397–423.

Fujii, E., M. Khaled, J. Mak, et al. (1985). The exportability of hotel occupancy and other tourist taxes. *National Tax Journal 38*(2), 169–177.

General Statistic Office (1995). *Statistical Yearbooks 1975–94*. Hanoi: Hanoi: Statistical Publishing House.

General Statistic Office (2000). *Statistical Yearbook of Vietnam 1999*. Hanoi: Hanoi: Statistical Publishing House.

General Statistic Office (2004). *Vietnam Household Living Standards Survey 2004*. Hanoi.

General Statistic Office (2006a). Vietnam household living standard survey.

General Statistic Office (2006b). *Vietnam Household Living Standards Survey 2006*. Hanoi.

General Statistic Office (2008a). Exports in march and the first quarter of 2008 (xuat khau hang hoa thang 3 va quy 1 nam 2008). Downloaded from http://www.gso.gov.vn in August 25, 2012.

General Statistic Office (2008b). *Statistical Data of Vietnam, 1986–2006*. Hanoi: Hanoi: Statistical Publishing House.

General Statistic Office (2009). *Statistical Yearbook of Vietnam 2008*. Hanoi: Hanoi: Statistical Publishing House.

Ghosh, M. and J. Whalley (2004). Are price controls necessarily bad? the case of rice in vietnam. *Journal of Development Economics 73*(1), 215–232.

Glewwe, P., M. Gragnolati, and H. Zaman (2002). Who gained from vietnam's boom in the 1990s? *Economic Development and Cultural Change 50*(4), 773–792.

Government of Vietnam (2008). Hanoi: 12th National Assembly, 4th meeting, 13 November 2008, Report No. 177/BC-CP., Title = Report and response to questions by Prime Minister Nguyen Tan Dung.

Government of Vietnam (2010). Decree on rice exporting business (Nghi dinh ve kinh doanh xuat khau gao). Government of Vietnam: Decree No. 109/2010/ND-CP.

Grafton, Q., T. Kompas, and P. Owen (2007). Bridging the barriers: Knowledge connections, productivity and capital accumulation. *Journal of Productivity Analysis 28*(3), 219–231.

Green, E. J. and W. E. Strawderman (1991). A James-Stein Type Estimator for Combining Unbiased and Possibly Biased Estimators. *Journal of the American Statistical Association 86*(416), 1001–06.

Han, C. and P. Schmidt (2001). The Asymptotic Distribution of the Instrumental Variable Estimators when the Instruments are not Correlated with the Regressors. *Economics Letters 74*(1), 61–66.

Han, T. and T. Wahl (1998). China's rural household demand for fruit and vegetables. *Journal of Agricultural and Applied Economics 30*, 141–150.

Hancock, D. (1985). The financial firm: Production with monetary and nonmonetary goods. *The Journal of Political Economy 93*(5), 859–880.

Hansen, H. and T. Nguyen (Eds.) (2007). *Market Policy and Poverty Reduction in Vietnam.* Vietnam Culture and Information Publishing House.

Hausman, J. and W. Taylor (1981). Panel Data and Unobservable Individual Effects. *Econometrica 49*(6), 1377–98.

Hausman, J. and W. Taylor (1983). Identification in linear simultaneous equations models with covariance restrictions: an instrumental variables interpretation. *Econometrica: Journal of the Econometric Society 51*(5), 1527–1549.

Headey, D. and S. Fan (2008). Anatomy of a crisis: the causes and consequences of surging food prices. *Agricultural Economics 39*, 375–391.

Hoang, H., G. Pham, M. Tran, and H. Hansen (2007). Ethnicity and poverty reduction. In T. Nguyen and H. Hansen (Eds.), *Market Policy and Poverty Reduction in Vietnam.* National Political Publisher.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science 14*(4), 382–401.

Horridge, M. (2003). Orani-g: a generic single-country computable general equilibrium model. Centre of Policy StudiesCOPS. Monash University. Australia.

Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge University Press.

Huynh, T. B., B. H. Duong, and T. C. Bui (2002). *Indigenous Peoples/Ethnic Minorities and Poverty Reduction.* Asian Development Bank.

Ivanic, M. and W. Martin (2008). Implications of higher global food prices for poverty in low-income countries1. *Agricultural economics 39*, 405–416.

James, W. and C. Stein (1961). Estimation with quadratic loss. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Volume 1, pp. 361–379. Berkeley, CA; University of California Press.

Judge, G. and R. Mittelhammer (2004). A semiparametric basis for combining estimation problems under quadratic loss. *Journal of the American Statistical Association 99*(466), 479–87.

Kalirajan, K. and M. Obwona (1994). Frontier production function: The stochastic coefficients approach. *Oxford Bulletin of Economics and Statistics 56*(1), 87–96.

Khiem, N. and P. Pingali (1995). Supply responses of rice and three food crops in vietnam. In G. Denning and V.-T. Xuan (Eds.), *Vietnam and IRRI: A Partnership in Rice Research*, pp. 275–290. International Rice Research Institute, Philippines and Ministry of Agriculture and Food Industry, Hanoi, Vietnam.

Kodde, D. and F. Palm (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: journal of the Econometric Society 54*(5), 1243–1248.

Kompas, T. (2004). Market reform, productivity and efficiency in vietnamese rice production. Working Papers, Crawford School of Economics and Government, Australian National University, Canberra.

Kompas, T., N. T. Che, H. T. M. Nguyen, and H. Q. Nguyen (2012). Productivity, net returns and efficiency: Land and market reform in vietnamese rice production. *Land Economics 88*(3), 478–495.

Krogstrup, S. and S. Wälti (2008). Do Fiscal Rules Cause Budgetary Outcomes? *Public Choice 136*(1), 123–138.

Lazear, E. and R. Michael (1988). *Allocation of Income within the Household.* University of Chicago Press.

Le, C. (2008). An empirical study of food demand in vietnam. *ASEAN Economic Bulletin 25*(3), 283–292.

Lim, M. (2006). *An assessment of the impact of the North American Free Trade Agreement (NAFTA) on the US textile industrys production activities: Qualitative and quantitative approaches.* Ph. D. thesis, Virginia Polytechnic Institute and State University.

Lipton, M. and M. Ravallion (1995). Poverty and policy. *Handbook of development economics 3*, 2551–2657.

Luu, H. T. D. (2003). The organization of the liberalized rice market in vietnam. Unpublished PhD Thesis, Groningen University, the Netherlands.

Luu, V. (2010). *Mot so diem nong chinh tri xa hoi dien hinh tai cac vung da dan toc o mien nui trong nhung nam gan day: Hien trang, van de, cac bai hoc kinh nghiem trong xu ly tinh huong (Some critical socio-political areas in mountainous multi-ethnic regions in recent years: situation, issues and lessons learnt in handling situations).* The national politics publisher (Nha xuat ban chinh tri quoc gia).

MAFI (1991). A report on mechanization in agriculture 1986-90 and the strategy for the next five years. Hanoi: Unpublished manuscript.

MAFI (1994). *Annual Reports, 1975–94*. Hanoi: Hanoi.

Marsh, S. and T. MacAulay (2002). Land reform and the development of commercial agriculture in vietnam: policy and issues. *Australian Agribusiness Review 10*, 1–19.

Marsh, S. P., V. P. Hung, D. T. Nguyen, and G. T. MacAulay (2006). Farm size change and the market for agricultural land use rights in vietnam since 1993. In S. Marsh, G. T. MacAulay, and V. P. Hung (Eds.), *Agricultural Development and Land Policy in Vietnam*. Canberra: Australian Centre for International Agricultural Research, Canberra.

McMillan, J., J. Whalley, and L. Zhu (1989). The impact of china's economic reforms on agricultural productivity growth. *The Journal of Political Economy 97*(4), 781–807.

Meeusen, W. and J. van den Broeck (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review 18*(2), 435–444.

Ministry of Education and Training, UNICEF and UNESCO (2008). *The Transition of Ethnic Minority Girls in Vietnam from Primary to Secondary Education*. Hanoi: UN.

Ministry of Labour, Invalids & Social Affairs and UNDP (2004). *Taking Stock, Planning Ahead: Evaluation of the National Targeted Programme on Hunger Eradication and Poverty Reduction and Program 135*. UNDP. Hanoi: UNDP.

Minot, N. and F. Goletti (1998). Export liberalization and household welfare: the case of rice in vietnam. *American Journal of Agricultural Economics 80*(4), 738–749.

Minot, N. and F. Goletti (2000). *Rice market liberalization and poverty in Viet Nam*. Number 114. International Food Policy Research Institute.

Mittelhammer, R. C. and G. G. Judge (2005). Combining estimators to improve structural model estimation and inference under quadratic loss. *Journal of Econometrics 128*(1), 1–29.

Mitze, T. (2009). Endogeneity in panel data models with time-varying and time-fixed regressors: to IV or not IV? Ruhr Economic Paper No. 83.

Mu, R. and D. Van De Walle (2007). Rural roads and poor area development in Vietnam. World Bank Policy Research Working Paper No 4340.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica 46*(1), 69–85.

MWR (1994). *Statistical Data, Water Sources Sector of Vietnam 1986–90.* Hanoi: The Gioi Publishing House.

Nguyen, H., T. Kompas, T. Breusch, and M. Ward (2012). Language, mixed communes and infrastructure: Sources of inequality and ethnic minorities in vietnam. Crawford School of Public Policy, The Australian National University, Crawford School Research Paper No. 12-07.

Nguyen, K. (1995). Vietnam: Agricultural Policy Reforms and Microeconomic Environment. Paper presented at the Vietnam Update, Australian National University, Canberra.

Nguyen, L. H., D. Brennan, Q. N. Nguyen, and N. Q. Nguyen (2009). Demand for food in Vietnam. Institute of Policy and Strategy for Agriculture and Rural Development. Draft August 2009.

Nguyen, T. T. P. and B. Baulch (2007). A Review of Ethnic Minority Policies and Programs in Vietnam. A product of the IDS-UoS-CAF Project on Ethnic Minority Development in Vietnam financed by the UK Economic and Social Research Council (RES-167-25-0157) and the Department for International Development.

Nielsen, C. (2003). Vietnam's rice policy: recent reforms and future opportunities. *Asian economic journal 17*(1), 1–26.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review 14*(3), 693–709.

Pandey, S., N. Khiem, H. Waibel, and T. Thien (2006). *Upland rice, household food security, and commercialization of upland agriculture in Vietnam.* International Rice Research Institute.

Parker, S., L. Rubalcava, and G. Teruel (2005). Schooling inequality and language barriers. *Economic Development and Cultural Change 54*(1), 71–94.

Patrinos, H., E. Velez, and G. Psacharopoulos (1994). Language, education, and earnings in asuncion, paraguay. *The Journal of Developing Areas 29*(1), 57–68.

Pham, T. H., D. T. Le, and V. C. Nguyen (2011). Poverty of Ethnic Minorities in Viet Nam: Situation and Challenges in Programme 135 Phase II Communes, 2006-07.

Phap Luat (2010). Giao su Vo Tong Xuan: Doc quyen, gia gao se con rot the tham (Professor Vo Tong Xuan: Monopoly, the price of rice will continue to fall dramatically). Phap Luat online newspaper dated 8 October, 2010. Downloaded from http://phapluattp.vn/201007071139449p1014c1070/giao-su-vo-tong-xuan-doc-quyen-gia-gao-se-con-rot-the-tham.htm in August, 25th, 2012.

Phien, D. V. (2001). Mechanization of Rice Production in Vietnam. Paper presented at the International Workshop on Agricultural Mechanization: Issues and Priorities in the New Development Stage, Hanoi.

Phung, T. D. and H. Waibel (2010). Food price crisis, poverty and welfare in vietnam: An ex-post decomposition analysis. Working Paper. Leibniz Universitt Hannover, Institute of Development and Agricultural Economics, Germany. December, 2010.

Plümper, T. and V. E. Troeger (2007a). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis 15*(2), 124–39.

Plümper, T. and V. E. Troeger (2007b). xtfevd.ado version 2.00 beta. Accessed from http://www.polsci.org/pluemper/xtfevd.ado.

Ravallion, M. and D. Van de Walle (1991). The impact on poverty of food pricing reforms: a welfare analysis for indonesia. *Journal of Policy Modeling 13*(2), 281–299.

Ravallion, M. and D. Van de Walle (2008). *Land in transition: Reform and poverty in rural Vietnam.* Washington, D.C.: World Bank.

Sawa, T. (1973). The mean square error of a combined estimator and numerical comparison with the TSLS estimator. *Journal of Econometrics 1*(2), 115–32.

Schmidt, P. and C. Knox Lovell (1979). Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *Journal of econometrics 9*(3), 343–366.

SDAFF (1991). Statistical data of agriculture, forestry and fisheries in vietnam for 35 years: 1956-90. Joint documentation by SDAFF (GSO) and the Institute of Planning and Construction for Agriculture.

SDAFF (1992). *Statistical Data of Vietnam: Agriculture, Forestry and Fisheries 1976–1991*. Hanoi: Hanoi: Statistical Publishing House.

SDAFF (2001). *Statistical Data of Vietnam: Agriculture, Forestry and Fisheries 1975–2000*. Hanoi: Hanoi: Statistical Publishing House.

SDAFF (2006). *Statistics of National Census of Agriculture, Forestry and Fisheries 2006*. Hanoi: Hanoi: Statistical Publishing House.

Shea, J. (1993). Do supply curves slope up? *The Quarterly Journal of Economics 108*(1), 1–32.

Shigetomi, S., K. Kubo, and K. Tsukada (2011). Rice trade in the 2008 food crisis. In S. Shigetomi, K. Kubo, and K. Tsukada (Eds.), *The World Food Crisis and the Strategies of Asian Rice Exporters*, pp. 1–21.

Sicular, T. (1988). Plan and market in chinas agricultural commerce. *Journal of Political Economy 96*(2), 283–307.

SPC (1995). Principles of changing agricultural structures in vietnam rural areas. Hanoi: Unpublished manuscript.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In D. W. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press.

Swinkels, R. and . Turk, C. (2006). Explaining ethnic minority poverty in vietnam: a summary of recent trends and current challenges. Draft Background Paper for Committee of Ethnic Minorities/Ministry of Planning and Investment Meeting on Ethnic Minority Report, 28 September 2006.

Tang, A. M. (1980). Food and agriculture in china: Trends and projections, 1952–77 and 2000. In A. M. Tang and B. Stone (Eds.), *Food Production in the Peoples Republic of China*. Washington, D.C.: International Food Policy Research Institute.

Thanh, T. (2008). Land consolidation: Strategy for the Future. *The World and Vietnam Report*. Hanoi: Ministry of Foreign Affairs.

Timmer, C. (2008). Causes of high food prices. Asian Development Bank Economics Working Paper 128.

Timmer, C. (2010). Reflections on food crises past. *Food Policy 35*(1), 1–11.

Timmer, C. (2012). Behavioral dimensions of food security. *Proceedings of the National Academy of Sciences 109*(31), 12315–12320.

Tran, T. T. T. (2004). Vietnam's rural transformation: Information, knowledge and diversification. In D. McCargo (Ed.), *Rethinking Viet Nam*. London: Routledge Curzon.

Tsukada, K. (2011). Vietnam: Food security in a rice-exporting country. In S. Shigetomi, K. Kubo, and K. Tsukada (Eds.), *The World Food Crisis and the Strategies of Asian Rice Exporters*, pp. 53–72.

Tuoi Tre (2008). Lua gao dang bi lam gia! (Rice and paddy is being overpriced!). Tuoi Tre online newspaper dated 27 April, 2008. Downloaded from http://tuoitre.vn/Chinh-tri-Xa-hoi/254768/Xu-ly-ngay-hanh-vi-gam-hang-dau-co-gao.html in August, 25th, 2012.

UNCTAD (2007). ATPSM database. United Nations Conference on Trade and Development. Downloaded from http://r0.unctad.org/ditc/tab/atpsm-download.shtm.

Van de Walle, D. and D. Gunewardena (2001). Sources of ethnic inequality in viet nam. *Journal of Development Economics 65*(1), 177–207.

Van Hung, P., T. MacAulay, and S. Marsh (2007). The economics of land fragmentation in the north of vietnam. *Australian Journal of Agricultural and Resource Economics 51*(2), 195–211.

Vasavakul, T. (2003). Language policy and ethnic relations in vietnam. In E. M. Brown and S. Ganguly (Eds.), *Fighting Words: Language Policy and Ethnic Relations in Asia*. Cambridge: MIT Press.

Vietnam Academy of Social Sciences (2009). *Participatory Poverty Assessment 2008: Synthesis Report*. Hanoi: The Gioi Publisher.

Vietnam Academy of Social Sciences (2011). *Poverty Reduction in Vietnam: Achievements and Challenges*. Hanoi: The Gioi Publisher.

Vo, X. T. (1995). *Rice production, Agricultural Research, and the Environment: Vietnam's Rural Transformation*. Boulder, Colorado: Westview Press.

Von Braun, J. (2008). *Food and financial crises: Implications for agriculture and the poor.*, Volume 20. International Food Policy Research Institute.

Vu, L. and P. Glewwe (2011). Impacts of rising food prices on poverty and welfare in vietnam. *Journal of Agricultural and Resource Economics 36*(1), 14.

Warr, P. (2005). Food policy and poverty in indonesia: a general equilibrium analysis. *Australian Journal of Agricultural and Resource Economics 49*(4), 429–451.

Warr, P. (2008). World food prices and poverty incidence in a food exporting country: a multihousehold general equilibrium analysis for thailand. *Agricultural Economics 39*, 525–537.

Wong, K.-f. (1997). Effects on inference of pretesting the exogeneity of a regressor. *Economics Letters 56*(3), 267–71.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press.

World Bank (1994). *Vietnam: Toward the Market Oriented Economy.* Hanoi: The National Political Publishing House.

World Bank (2003a). *Vietnam: Delivering on its Promise.* Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.

World Bank (2003b). Implementation Completion Report on a Credit in the Amount of SDR 49.6 million (US $70 Million Equivalent) to the Socialist Republic of Vietnam for a VN-Primary Education Project. Report No 25967. Washington D.C.: The World Bank.

World Bank (2004). Vietnam Reading and Mathematics Assessment Study. World Bank Human Development Sector Report. Washington D.C.: World Bank.

World Bank (2005). Business. Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.

World Bank (2007). Social Protection. Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.

World Bank (2009a). *Country Social Analysis: Ethnicity and Development in Vietnam.* Washington D.C.: World Bank.

World Bank (2009b). *Capital Matters.* Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.

World Bank (2009c). Modern institutions. Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.

World Bank (2011). Market economy for a middle-income vietnam. Joint Donor Report to the Vietnam Consultative Group Meeting. Hanoi: World Bank.