
RESEARCH PROPOSAL

Presented to

PEP Network

By

Pham Thai Hung

&

Bui Anh Tuan, Dao Le Thanh

VIETNAM

First draft: 30 November 2006

Abstract: This study bases on the notion that promoting a buoyant rural nonfarm sector and its sectoral linkage with agriculture is the key to rural development and poverty reduction. It will use the Vietnamese household surveys data over the 1993-2004 period to investigate (i) the contribution of the nonfarm income to overall rural poverty and income inequality; (ii) the determinants of nonfarm diversification and its dynamics; (iii) to what extent nonfarm diversification affects rural household welfare, and how the poor and the non-poor benefit differently from nonfarm diversification. Findings from the empirical analysis will be discussed and enhanced through a policy dialogue with key agencies before formulating policy suggestions. The research is expected to help better understanding on the importance of nonfarm sector in rural poverty reduction, which is also understood to be an interest for policy makers and donors.

Acknowledgements: The authors are grateful to the reviewer(s) for useful comments on the first draft of this proposal; and to Swapna Mukhopadhyay, Lili Lu, and Steve Younger and other participants at the 6th PEP General Meeting in Lima, Peru on the comments on the first revised versions. In addition, we would like to thank Abdelkrim Araar, Aissatou Diop for their support in finding relevant documents for this revision; Nguyen Hai Dong, Nguyen Quang Ngoc, and Do Xuan Thong for sharing their experience and policy concerns with respect to the focus of this study. Usual disclaimer applies.
1. Main research questions and core research objectives

As appeared in the title, the main research question of this proposal is that: “Is Nonfarm Diversification a Way Out of Poverty for Rural Households?”. In particular, we will investigate the following research questions:

− How important is nonfarm income for rural households? To what extent the nonfarm income source has contributed to overall rural poverty and income inequality?
− What are determinants of nonfarm diversification by rural households?
− What are the determinants of nonfarm income? To what extent does nonfarm diversification contribute to improve welfare of rural households? Whether the poor and the non-poor benefit differently from the RNFS?
− What policy suggestions can be drawn for promoting nonfarm employment as a measure to the reduction of rural poverty and inequality?

These questions will be researched under the context of the Doi moi (renovation) process in Vietnam, using the four Vietnamese Household Living Standard Surveys available in the period 1993-2004.

Given the above research questions, this study aims at the following objectives:

− For academia: to contribute to the growing literature on rural nonfarm sector by providing empirical evidence on the causal relationship between nonfarm diversification and household welfare.
− For policy makers: to provide sound policy suggestions for promoting the nonfarm sector, and rural poverty reduction on the basis of the empirical findings.
− For capacity-building: to build up a publication track record for the research team on the research area (one or two papers are targeted to high-ranking scholarly journals).

2. Scientific contribution of the research in the literature and knowledge gaps

The impact of nonfarm diversification on household welfare is a complicated issue, both theoretically and empirically. While participating in nonfarm activities apparently contributes to total household income, there has been a debate on the interaction between nonfarm diversification and poverty reduction. This study is proposed to contribute to this literature.
The recent literature on rural nonfarm sector (RNFS) in developing countries tends to suggest a mixed effect of nonfarm diversification on household welfare. Lanjouw and Lanjouw (1995) consider RNFSs a combination of both productive and non-productive activities. While the former is likely to considerably raise living standards of rural households, the latter is described as ‘residual’ activities by rural households in response to shortfalls of income. In this regard, the welfare effect of nonfarm diversification depends on whether rural households are in a ‘pull’ or ‘push’ scenario – using Hart’s (1994) terminology. Some rural households may be ‘pushed’ into nonfarm activities in their struggle to survive, while others may be ‘pulled’ into them by their desire to accumulate. As the ‘pushed’ scenario is usually referred to poor households and the ‘pulled’ is more likely associated with the non-poor, the welfare effect of nonfarm diversification on rural poverty in general is not unequivocal. Ellis (1998) supports this argument and urges that nonfarm participation may be associated with success at achieving livelihood security under improving economic conditions as well as with livelihood distress in deteriorating conditions. According to Von Braun and Pandya-Lorch (1991) rural households seek nonfarm activities either for ‘good’ or for ‘bad’ reasons. While the latter refers to the pressure on the poor to participate in the RNFS as a coping strategy, the former implies the attraction of the nonfarm sector to the better-off.

Given this, the welfare effect of nonfarm diversification largely depends on supply-side availability and dynamics of RNFSs, and household capacity to participate and take advantages of nonfarm opportunities. Nonfarm diversification is more welfare-enhancing when it occurs in a dynamic rural economic base, with improving infrastructure conditions, and/or when households have certain capacity (i.e. human capital, lands and other assets) to undertake investment into such opportunities. Therefore, the effect of nonfarm diversification on household welfare depends on specific context of research and remains largely an empirical question. In this context, there have been a growing number of empirical studies on this issue. In Japan, Taiwan, and South Korea, the poorer/landless households experienced a higher percentage of income from nonfarm activities, and this suggests an equalizing influence and poverty alleviation role of the RNFS (Lanjouw and Lanjouw, 2001). Ravallion and Datt (2002) find that farm yield and nonfarm output are all associated with poverty reduction in different states of India. In Berdegue et al. (2001) and Lanjouw (2001), the poor are found to be engaged in ‘last resort’ nonfarm activities, while the non-poor are active in productive nonfarm activities in El Salvador and Chile, respectively. By reviewing 18 field studies, Reardon (1997) shows that the share of nonfarm income in total income is two twice higher in upper third households compared to lower third households. In general, the existing studies reveal either a U-shaped or a negatively-sloped relationship between nonfarm income and total household income or assets.
The evidence above has been obtained mainly on the basis of descriptive analysis. There are few studies that tackle the relationship between nonfarm diversification and household welfare by using econometric models. The endogeneity concern of diversification to poverty is probably the main difficulty in establishing a causal relationship between nonfarm diversification and household welfare. As noted in Pham (2006), most of the current empirical studies on the RNFS focus either on the probability of nonfarm diversification or the determinants of nonfarm incomes or both. To our knowledge, there are a few exceptions that formally deal with the relationship between nonfarm diversification and household welfare. These include Reardon et al. (1992), Lanjouw (1998), Van de Walle and Cratty (2003), Dabalen et al. (2004), De Janvry, Sadoulet, and Zhu (2005), Bezemer et al. (2005), and Jonasson (2005). These studies are briefly reviewed below.

Reardon et al. (1992) employ a recursive system to examine the interaction between nonfarm diversification, household income, and consumption expenditures in Burkina Faso and reveal a positive impact of nonfarm diversification on household income and food consumption. In the case of Ecuador, Lanjouw (1998) proposes a simple simulation that involves estimating an earnings regression over the whole population of wage-earners and using the estimates to predict the average earnings of the poor. Lanjouw found that a shift of the poor out of the traditional sector into non-agricultural activities would imply a rise in the average income. By estimating the individual earnings equation and household expenditures, Jonasson (2005) reports a better earnings potential for rural households in the RNFS in Peru. De Janvry et al. (2005) examine the earnings potential in the RNFS more thoroughly by simulating a counterfactual of what the welfare outcomes (in terms of household incomes, poverty, and inequality) would be in the absence of nonfarm activities. De Janvry et al. then reveal that without nonfarm income sources, rural poverty and income inequality would be much higher in Hubei province of China. Bezemer et al. (2005) introduce a departure from the classical regression approaches to apply a Bayesian stochastic frontier approach (though the OLS is also used) in estimating technical efficiency of households who involved in both farming and nonfarm activities in Georgia. The results demonstrate that nonfarm diversification has contributed to higher technical efficiency in agriculture and higher income. In the case of Vietnam, Van de Walle and Cratty (2003) provide some insights on the relationship between nonfarm activities and rural poverty by using a ‘common causation’ method. This involves identifying exogenous variables having the same sign in both welfare and diversification regressions. Although this study points out variables that jointly influence both living standards and nonfarm diversification, it does not offer conclusive evidence on the causality. With efforts to tackle the same issue, Dabalen et al. (2004) use a semi-parametric approach, the Propensity Score Matching, to examine the welfare impact of nonfarm diversification in rural Rwanda. By
comparing earnings of different household groups, they generally conclude that participating in nonfarm activities produces a positive impact on household welfare.

Though these studies have demonstrated the potential of the RNFS in contributing to incomes of rural households, and thus rural poverty reduction, the conclusions are far from conclusive, and further empirical evidence is needed regarding the causal relationship between nonfarm diversification and household welfare. To this knowledge gap, does this study aim at providing insights on the welfare effect of nonfarm diversification in the context of rural Vietnam in the period 1993-2004. The case of Vietnam is selected for a number of reasons. Firstly, Vietnam has been commonly considered as one of the top two or three performers in the developing world over the past two decades. Within just ten-year time the widespread poverty was nearly halved (from 58% in 1993 to 29% in 2002)\(^1\) and this impressive poverty reduction was achieved without any significant increases in the overall income inequality (see Glewwe et al. 2004). Though the impacts of Doi moi on household welfares have been well documented, how the RNFS contributes to the country’s experience in poverty reduction has been received little attention (see below for further discussions).

More importantly, the case of rural Vietnam during the past two decades is characterized by a vigorous rural transformation process in which rural households have been considerably diversified out of agriculture. In terms of individual diversification, the proportion of rural workers who were employed in the RNFS as their main jobs has increased from 22% to nearly 36% in between 1993 and 2002. During the same period, the incidence of farm-only households fell by 20% (see Table A1). Given this, our calculations from the recent household surveys (see below) reveal that nonfarm income has become an increasingly important source of income for rural households (see Table A2). With an average share of 38 percent during the period 1993-2002, the share of nonfarm income in Vietnam is as high as those reported in Africa and Latin America, and higher than the average level of the other Asian countries (e.g. China, India, Philippines, and Pakistan) (see Table A3). With this development of the RNFS, it is reasonable to expect a strong and positive contribution of nonfarm diversification to rural poverty reduction. This expected correlation is evident in Figure 1A. Based on the kernel densities of the real expenditures, the farm-only households are among the poorest, the diversified households ranked second, and the nonfarm-only households are revealed as the richest (see Figure A1). Last but not least, it is very fortunate that there has been a series of high quality surveys on household living standards that spread over Doi moi. This series has been intensively used in most of the quantitative research on Vietnam since the early 1990s and this study will also employ this valuable dataset for its empirical analysis.

\(^1\) Calculations from the VLSS 1992/93 and VHLSS 2002.
3. **Policy relevance**

There are generally two approaches toward rural poverty reduction underlying an array of initiatives undertaken by governments, international development agencies, and NGOs. The traditional approach has been through redistributive land reforms and integrated rural development programmes to increase the productivity in agriculture of the assets controlled by the rural poor. However, this approach has had limited success. After many decades of focusing on agriculture as the core of rural development policies, it is now clear that agriculture is not enough for sustainable increases in incomes (World Bank, 1997). The failure of this integrated rural development approach has led to the development of the recent alternative, which is characterized by decentralization, local institutions, participations, and a demand-driven approach to the allocation of public resources. This new approach recognizes the multiplicity of household income sources in a particular regional setting (De Janvry and Sadoulet, 2001). As a result, promoting sectoral linkages between agriculture and the RNFS, generation of nonfarm employment opportunities, and efforts to enhance access for the rural poor to these income sources is one of the key features of this new approach.

The situation of rural development in Vietnam during the *Doi moi* mirrors the above switch in the rural development approaches. Since the early 1990s, agriculture has been central to the country’s impressive growth and poverty reduction. The land reform, trade liberalization, and promotion of the household sector were crucial for a robust agricultural growth and hence incomes of rural households (Benjamin and Brandt, 2002; World Bank, 1998, 2006a). However, it has recently become a concern that agriculture will not be sufficient to sustain the rapidly growing population (Van de Walle and Cratty, 2003; World Bank, 2006a,b). In this context, promoting nonfarm diversification is now recognized to be critical to the country’s future growth and poverty reduction. This switch toward the new rural development approach has been reflected in the Five Year Plan for Socio-Economic Development 2001-2005, which states the objective of promoting processing industries, rural industries and services to generate employment and farmers’ income (MPI, 2003). This priority is also emphasized on the Comprehensive Poverty Reduction and Growth Strategy (CPRGS) (SRV, 2003). Despite a plethora of policy interventions to assist rural transformation toward a more diversified and dynamic structure, the recent Five Year Plan for Socio-Economic Development 2006-2010 reveals that this transformation has been slow and lack of sustainability (MPI, 2006). A continuing priority toward the employment structural shift out of agriculture remains an important goal of this five year plan.

Despite the importance attached to nonfarm diversification, our understanding on the RNFS in Vietnam is currently limited. To our knowledge, Van de Walle and Cratty (2003), Hoang,
Dang and Tacoli (2005), Minot, Epprecht, Tran and Le (2006), and Pham (2006) are few existing studies on the RNFS in Vietnam. Van de Walle and Cratty (2003) using the VLSSs reveal that the incidence of farm-only household has decreased from 75% to 52% from 1993 to 1998. Pham (2006) using individuals as unit of observation report that the share of the RNFS in rural employment has increased from 22% in 1993 to 32% in 2002. Hoang et al. (2005) collect information from two villages in the Red River Delta and reveal an important role of nonfarm activities in poverty reduction. Minot et al. (2006) focus on income diversification by households the Northern Uplands with a greater emphasis on agricultural diversification but provide little on nonfarm diversification in this region. While these studies has a common message of showing an increasingly important of the RNFS in rural transformation and poverty reduction, the effect of nonfarm diversification on household welfare and rural poverty has not been examined thoroughly and remains an under-researched question.

This limited understanding of the contribution by the RNFS as a likelihood strategy for rural households and in rural poverty reduction is re-affirmed by our personal communication with some policy makers (see sub-section Policy dialogue for more details). For instance, Mr. Nguyen Hai Dong (Head, Department of Labour and Employment Policies, Ministry of Labour, War Invalids, and Social Affairs - MOLISA) revealed that the policies on rural labour market are generally under a broad umbrella of ‘rural industrialization and rural modernization’ strategy endorsed by the Communist Party’s IX National Congress. Despite the political will and support, there are lack of specific policy attention and measures to promote the RNFS. He also emphasized a need for better understanding of the evolution of the RNFS, how this sector contributes to incomes of rural households, and thus rural poverty reduction. Mr. Nguyen Quang Ngoc (Cooperation Development, Embassy of Sweden in Hanoi) shared his experience in the Sida-funded programmes in administrative reform and poverty reduction in Quang Tri province. One of the lessons drawn from these programmes is the importance of nonfarm income-generating opportunities in improving living standards for the rural households, especially for ethnic minority groups. However, as Mr. Do Xuan Thong (who is in charge of managing EU-supported projects at the Ministry of Planning and Investment - MPI) argued, there has been a few rural development projects with components that target at promoting nonfarm activities.

In this context, this study will provide an input for relevant stakeholders, including government officials and donors, regarding the contribution of the RNFS to overall rural poverty and income inequality, determinants of nonfarm diversification, and its effect on rural household

---

2 These communications were made in response to the suggestion by the reviewer(s) on the earlier version of this proposal. The first revision coincided with the Tet (Lunar Calendar New Year) holiday and its aftermath in Vietnam. As this is arguably the busiest period during the year, meeting policy makers was extremely difficult. These contacts were thus made through telephones between Bui Anh Tuan (team member) and these stakeholders.
welfare over the period 1993-2004. The expected research output is firstly to find out the relative importance of nonfarm income and its contribution to overall rural poverty and income inequality (*question 1*). For policy-makers who usually prefer simple and meaningful figures, the output from this question will be useful for a better understanding of how important the RNFS is and why this sector needs to be promoted. Once the importance and contribution of the RNFS in rural poverty reduction is recognized, the study will then inform the determinants of non-farm diversification (*question 2*). This will provide the policy makers with some hints on what factors need to be considered in order to promote the RNFS. The study will also inform the determination process of nonfarm income, to what extent nonfarm diversification affect household welfare, and how the poor and the non-poor benefit differently from nonfarm diversification (*question 3*). With this question, we expect to inform the relevant authorities whether access to (and hence benefits from) nonfarm diversification is limited for the poor as it has been concerned in some developing countries (see Araujo, 2003). These are important issues to be considered when planning and implementing rural poverty reduction policies in Vietnam under the context of the new approach to rural development.

Are the findings and policy implications from this study applicable to other developing countries? To the authors’ knowledge, we believe the answer is ‘Yes’. Table A3 suggests the popularity of nonfarm activities as an important source of household incomes in many developing countries. Though the context varies, the literature on the RNFS has suggested a number of common factors affecting participation to and earnings from the RNFS across countries such as human capital, household physical assets, and access to both institutional and physical infrastructures (see Pham, 2006 for a review). Given this, it is reasonable to expect that the findings on (and implications) on the impacts of the RNFS in rural poverty in one country (Vietnam in this case) might also be relevant to the other developing countries as well.

4. **Methodology**

The four research questions outlined above will be investigated using innovative quantitative methods drawn from the recent literatures on poverty analysis, likelihood studies, and labour market econometrics. Before making policy recommendations, the findings drawn from the empirical analysis will be discussed and verified through a number of semi-structured interviews with key stakeholders. We believe that this combination should shed lights on the welfare impact of nonfarm diversification on rural households and produce sound policy suggestions for the authorities in promoting a buoyant and pro-poor RNFS. To avoid having a complicated exposition of the methodology, we will describe the essences of the methods
proposed in the main text and present the technical details of the models in the Appendix 2. These methods are briefly described below.

**Question 1: How important is nonfarm income for rural households? To what extent has the nonfarm income source contributed to overall rural poverty and income inequality?**

A descriptive analysis approach and the Shapley approach will be applied to answer this question. With the simple descriptive analysis, a number of basic statistic summary indicators, concentration curves (such as kernel densities) will be produced from the dataset to portrait the evolution of the RNFS in the rural economy, and that of the nonfarm income source for rural households, and how the share of nonfarm income varies with the level of household welfare. The picture that emerges from this exercise will provide an overall context for further empirical analysis. The Shapley approach is then employed to decompose rural poverty (the FGT(a) indices for instance) and income inequality (the Gini coefficient or generalized entropy indices) by income components (see Shorrocks, 1999; Araar, 2002; Duclos and Araar, 2006; Araar, 2006). In this case, total income of rural households will be calculated from the four surveys and divided into four main components including (i) income from agriculture, (ii) nonfarm wage income, (iii) nonfarm self-employment income; and (iv) other income sources (see Table A2 for a brief explanation of these income sources). By applying the Shapley approach, this study will derive the contribution of each of the above income sources on overall rural poverty and income inequality (see Appendix 2 for more details).

**Question 2: What are determinants of nonfarm diversification by rural households?**

To investigate the determinants of nonfarm diversification by rural households, two models will be estimated, including (i) a multinomial logit model at the four ‘snapshots’ that coincide with the four surveys (VLSS 1992/93, VLSS 1997/98, VHLSS 2002, and VHLSS 2004); and (ii) a quantile regression approach. While the former is quite common in the literature (see Pham, 2006 for a survey), to our knowledge the use of the quantile regression approach to examine the determination of nonfarm participation has not been explored before.

In the MNL models, nonfarm diversification is measured by some polychotomous dependent variables. However, nonfarm activities, by nature, “[e]ncompass the full spectrum of economic activities which occur in rural areas but which are not directly associated with agriculture” (Lanjouw, 2001; p.529). Using discrete measures for nonfarm diversification may provide an incomplete picture of the determination of nonfarm activities due to the

---

3 The authors are grateful to the reviewer(s) for the suggestion (on the earlier version of this proposal) of using the Shapley approach.

4 The separation between nonfarm wage income and nonfarm self-employment income is necessary as wage earners might be systematically different self-employers, and it would certainly influence the earnings from nonfarm wage employment and nonfarm self-employment.
heterogeneity nature of these activities. Given this, we will measure nonfarm diversification as a continuous variable and adopt the quantile regression approach to investigate the determination of nonfarm diversification at different points of the conditional distribution of nonfarm activities. In addition, the quantile regression procedure is less sensitive to outliers and provides a more robust estimator in the face of departures from normality (Koenker, 2005; Koenker and Basset, 1978). In addition, Deaton (1997, pp.80-85) notes that quantile regression models may also possess better properties than other methods in the presence of heteroscedasticity.

In the quantile regression models, nonfarm diversification will be defined as the share of household members employed in the RNFS, which is a continuous variable. However, as this variable is only observed for households who have members employed in the RNFS, the sample is now truncated at zero. Therefore, the issue of potential selection bias need to be addressed before estimating the quantile regressions. In this case, the Heckman two-stage procedure can be used to obtain a correction bias correction term in the mean regression context. As argued in Buchinsky (2001), this term cannot be used for quantile regression approach because it does not represents the correction terms at different points of the conditional distribution that are needed in quantile regression models. In this research, we will apply a simple polynomial series expansion of the mean selection bias correction term as proposed in Buchinsky (1998, 2001). It is acknowledged that, in contrast to the mean regression case, this provides an inexact correction for selection bias. However, it circumvents the much trickier problem of identifying the quantile regression constant terms (see Pham and Reilly 2007 for a discussion).

Since Vietnam exhibits a great heterogeneity in resource endowments among regions, which range from highly fertile river deltas to less productive costal lowland, and infertile regions and erosion-prone hilly and mountainous hinterlands. This regional heterogeneity results in a very uneven development potential and hence uneven availability of nonfarm opportunities across the country. While estimating these two models, we will construct an index to show how far the mean probabilities by different regions differs from the weighted mean using the method popularized by Krueger and Summers (1988), and Zanchi (1998). This approach is common in the inter-industry literature but has not been applied to probability differentials in nonfarm diversification before. Since the index can be computed for each of the four years for which we have survey data, it allows the regional differentials in nonfarm diversification to be investigated by examining deviations of different regions from the national mean over time.

*Question 3: What are the determinants of nonfarm income? To what extent does nonfarm diversification contribute to improve welfare of rural households? Whether the poor and the non-poor benefit differently from the RNFS?*
There are three components of this question. In order to deal with the first component of the nonfarm income determination process, we will use both the mean and quantile regression approaches. Nonfarm income will be measured from wage income and nonfarm self-employment activities. Similar to the above, the issue of potential selection bias must be addressed before the income regression models can be estimated. For the mean regression, we will employ the Heckman two-stage procedure to correct for the problem of selectivity bias. In the quantile regression context, a similar version of the Buchinsky’s (1998) polynomial series expansion will be applied (see Pham and Reilly, 2007 for further details). The application of the quantile regression is proposed in conjunction with the mean regression to take into account the heterogeneity of nonfarm activities (as above).

For the second component, determining the causal relationship between nonfarm diversification and household welfare is constrained by the potential endogeneity. To investigate the extent that nonfarm diversification contributes to improve rural household welfare, we will apply (i) the two-stage least squares (TSLS) method; and (ii) the Propensity Score Matching (PSM). The former will be used primarily to examine the causal relationship between nonfarm diversification and household welfare. Meanwhile, the latter provides some straightforward estimates of the welfare gains from nonfarm diversification. In applying the TSLS method, the key is to find relevant instruments which are orthogonal to the error process in the structural equation of household welfare. Based on our knowledge of the survey data, it is likely to construct a number of instruments which affect the diversification decision-making process at the household level but not influence household expenditures (for instance, those reflect the availability of nonfarm opportunities at the community level).

Another approach to investigate the second research questions is to compare the differences between consumption expenditures of those who diversify into the RNFS and the ‘counterfactual’ levels that they forgone by such diversification. In this case, the Propensity Score Matching (PSM) method will be adopted. The idea is to create a group of diversified households (i.e. the treatment group) and a group of undiversified households (i.e. the control group) in a way that ensures that they are as similar as possible. This involves finding the same observable characteristics from the treatment group and the control group and then matching each diversified household with another undiversified household based on the similarity of these observable covariates. Rosenbaum and Rubin (1983) argue that this matching can be

---

5 As common in literatures on poverty analysis, household consumption expenditures will be used as a measure for household welfare.

6 The other alternative approach to overcome the potential endogeneity of nonfarm diversification is to assume a recursive system as in Reardon et al. (1992). Though we believe that the simple TSLS proposed in this study can be applied to examine the causality between nonfarm diversification and household welfares, this does not exclude the possibility of adopting a recursive system to investigate the welfare effect of nonfarm diversification.
easily made by using a single index, the propensity score, by assuming that conditional on the observables, there are no systematic differences between the treatment and control group.

In this study, applying the PSM approach is carried out in the two steps. The first step involves estimating the propensity score of a household being diversified in the RNFS. The second step is to calculate the welfare impact of nonfarm diversification (i.e. the average treatment effect on the treated) by matching the treatment (or the diversified households) with the control (or the undiversified households) based on the estimated propensity scores (see the technical appendix for more details). The PSM method uses observable characteristics to reduce the bias that is attributable to the unobservables, and as argued in Becker and Ichino (2002), the extent to which the bias is reduced depends on the quality of the observables. Given our knowledge of the data available from the four surveys, we believe that it is likely to construct a set of variables that best capture the observables that affect the probability of nonfarm diversification by rural households.

The literature reviewed earlier shows that the poor and the non-poor are likely to diversify in nonfarm sectors for different reasons and thus they might benefit differently from such diversification. To examine how do the poor and non-poor households benefit differently from the RNFS, these TSLS and PSM approaches will be carried out for the poor sub-sample and the non-poor sub-sample. Since a large number of households may be concentrated around the poverty line, we will thus implement these procedures using different adjustments. These adjustments include (i) using a poverty line 10% above (and below) the official poverty line; (ii) adjusting household consumption expenditure by adult equivalent units (see Pham et al. 2006 for more details).

**Question 4: What policy suggestions can be drawn for promoting nonfarm employment as a measure to the reduction of rural poverty and inequality?**

Given the findings from the empirical estimation, the final step is to formulate sound policy suggestions. This is our general impression that policy makers in Vietnam (or probably elsewhere) tend not to rely solely on ‘scientific evidence’ on their decision making processes. In order to make sound policy suggestions in this study, we will use our personal contacts with relevant policy makers and other appropriate stakeholders in Vietnam to establish around 7-10 semi-structured meetings (see dissemination strategy for more details). At these meetings, we will discuss our findings and intended policy suggestions to get their feedbacks before finalizing the policy suggestions in the final reports. By this interaction, we expect both to get

---

7 This separation will be made on the basic of the Chow test (or a likelihood ratio version of the Chow test) to investigate if it is statistically justifiable to separate between the poor and non-poor sub-samples.

8 See Saumier (2003) for an interesting and in-depth discussion on how policy making process is generally taken place and the impact of research on it in Vietnam’s context.
further input from them as well as to reinforce possible impacts of this study on policy formulation.

5. Data requirements and sources

Vietnam is fortunate to have a number of high quality household surveys. The first two Vietnam Living Standards Surveys (VLSSs) of 1992/93 and 1997/98 are multi-topic surveys patterned after the World Bank’s Living Standard Measurement Surveys with nationally representative samples of 4,800 and 6,000 households, respectively. These were followed in 2002 and 2004 by a new biennial household survey programme known as the Vietnam Household Living Standards Surveys (VHLSS), which uses a rotating core-and-module designed with an expanded sample size (30,000 households for the income-expenditure modules in the VHLSS 2002, and 9,000 households for the VHLSS 2004) which aims to produce statistics that are representative at the provincial level (the VHLSS 2006 is currently in the field). There is a panel of 4,300 households between the two VLSSs, and another panel of nearly 4,100 households between the two VHLSSs, but no panel linking the VLSSs and VHLSSs. Both the VLSSs and VHLSSs have clustered, stratified sampling designs.

Although the content of these surveys slightly varies across time, these are summarized by a household module and a community module. The household modules provide information on employment by individuals necessary to identify nonfarm diversification by households and members. In addition, the household modules cover a wide range of information from household size and composition, health, anthropometric measures of nutrition, education, housing characteristics, migration, nonfarm enterprises, agriculture, other income, expenditure, and food consumption, ownership of consumer durables, and savings and credit. The community modules of these surveys provide information on basic physical and demographic characteristics, general economic conditions and economic activities, physical infrastructure conditions and transportation, certain information on nonfarm employment opportunities, agricultural production, availability of credit and savings, as well as information on schooling and health status of households in each commune. Despite of certain modifications among the surveys, these data remain compatible and allow us to construct variables necessary for the econometric analysis proposed here. This study will use these four surveys to examine the research questions. The lead researcher has obtained the permission from the General Statistics Office (GSO) – the owner of this database – to use the VLSSs and the VHLSS 2002 for his PhD thesis and other related research projects. The research team will thus need to ask the GSO for the VHLSS 2004.
In addition, the study will make uses of the previous studies on the impact of the Doi moi on rural poverty reduction and labour market outcomes. Given the technical assistance by international donors over the past fifteen years, there has been now an intensive stock of empirical studies on rural poverty and poverty dynamics (see Pham et al. 2006; World Bank, 2006a, 2006b for references to this literature). Though the research theme of this project has not been studied in this literature, the previous studies on rural poverty in Vietnam would be a source of references.

6. Dissemination strategy

The dissemination strategy in this study includes a ‘policy dialogue’ component and a ‘normal’ dissemination component. The former is mainly proposed to stimulate a dialogue with relevant stakeholders about appropriate policy measures to combat rural poverty though promoting a buoyant RNFS. The latter is aimed at disseminating the research outputs by publications, seminars, and policy briefs.

Policy dialogue

An unusual component of this research project will be the use of the research team’s personal contacts to organize a number of semi-structured interviews with key stakeholders. In these small and face-to-face meetings, the research team will present and discuss the research findings with selected key agencies in Vietnam. These will include the Ministry of Labour, War Invalids, and Social Affairs (MOLISA), the Ministry of Agricultural and Rural Development (MARD), Ministry of Planning and Investment (MPI), National Target Programme for Hunger Eradication, Poverty Reduction and Job Creation (HEPRJC), the Central Institute for Economic Management (CIEM), and the Institute of Economics at the Vietnamese Academy of Social Sciences (VASS). In addition, these meetings will be held between the researchers and some leading international donors and NGOs with an interest in rural development in Vietnam, including Sida, DfID, World Bank country office, plus the International Support Group (a MARD-host donor forum on rural development - ISG).

It is important to note that this policy dialogue should not be over-emphasized within the scope and resources of the research team and this project. Instead, the research team expects to have from 7-10 interviews with the above institutions.9 Through these interviews, it is hoped that a

---

9 The researchers have made some contacts with a numbers of ministries officials and international donors in Vietnam. Tentatively, the following persons will be willing to participate in this study’s policy dialogue: (i) at ministries: Mr. Nguyen Dai Dong (Head of Department of Labour and Employment Policies, MOLISA), Mr. Nguyen Huu Hue, (Head of Labour Relation Division, Institute for Labour and Social Affair, MOLISA), Mr. Do Xuan Thong (Ministry of Planning...
shared understanding of nonfarm diversification and its importance in rural poverty reduction can be built-up and a dialogue about appropriate policy measures stimulated. The feedbacks from these meetings will be an additional input for the researchers to formulate policy recommendations. Due to the proliferation of project workshops and government/donor fatigue, it is believed that this strategy will be more effective than the big project inception and completion workshops that are common for research projects in Vietnam.

**Dissemination of research outputs**

Two types of written outputs will be produced by this project:

- The first will be a detailed technical report, which will form the basis of two papers to be targeted to refereed journals (such as World Development, Economic Development and Cultural Change, Journal of Comparative Economics). At this stage, we expect two papers, one focuses on the determination of nonfarm diversification, another on the welfare effects of nonfarm diversification. The former will be formulated largely on the basis of question 2, while the latter will be based on the findings of question 3. The results from remaining questions will be structured for these two papers as appropriate.

- The second type will be a short policy paper aimed at the Government and Communist Party Officials and donor agencies. This will also be written in Vietnamese and submitted to Vietnam’s Socio-Economic Development, the dual language journal of VASS, and *Tạp chí Kinh tế Phát triển* (i.e. *Journal of Economic Development* as translated to English), a Vietnamese language refereed journal of the National Economics University (NEU) – the host institution of the research team.

The content of these outputs will be presented at a numbers seminars host by the following, but not limited to, institutions (in addition to the seminars by the PEP Network):

- The Vietnam Development Forum, a joint-research project between the National Economics University (NEU) and Graduate Institute for Policy Research (GRIPS).

- Hopefully, the research output will be invited to present at the ISG Annual Plenary Meeting 2008 or as a side event during this Meeting (ISG Plenary Meetings attracted more than hundred participants from Vietnamese Ministries, donor community, international organizations and NGOs with interests in rural development).

- And hopefully, one or two other conferences/seminars abroad on the related issues.
In addition, the research outputs will be disseminated via websites to ensure the most convenient access by a wider public. This includes:

- The Vietnam Development Information Centres (http://www.vdic.org.vn), a multi-donor funded organisation, with centres in Hanoi, Ho Chi Minh City and Danang, and via the Vietnam NGO Resource Centre (located in Hanoi);

- The Vietnam Development Forum (http://www.vdf.org.vn); the Vietnam Economic Research Network (VERN) website (http://www.vern.org.vn/)

- Internationally, the project’s written outputs will be disseminated via the Poverty and Economic Policy (PEP) website (http://www.pep-net.org)

- Other websites of the NEU and also of the University of Sussex (where the lead research is currently in for his PhD program).

7. Short list of key references


Araujo, C. (2003), Non-agricultural employment growth and rural poverty reduction in Mexico during the 90s, mimeo, Department of Agricultural and Resource Economics, University of California, Berkeley.


- 18 -


8. List of team members’ prior training and experience in the issues and techniques involved.

The research team includes Pham Thai Hung (lead researcher), Bui Anh Tuan and Dao Le Thanh (team members):

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Gender</th>
<th>Prior training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pham Thai Hung</td>
<td>31</td>
<td>Male</td>
<td>MSc (SOAS, UK) Ph.D (Sussex, UK) (expected date: July 2007)</td>
</tr>
<tr>
<td>Bui Anh Tuan</td>
<td>43</td>
<td>Male</td>
<td>MA (Yonsei, Korea) Ph.D (NEU, Vietnam)</td>
</tr>
<tr>
<td>Dao Le Thanh</td>
<td>27</td>
<td>Female</td>
<td>MSc (Toulouse, France)</td>
</tr>
</tbody>
</table>

The lead researcher is a lecturer at NEU’s Department of Trade, who is currently on leave to pursue his Ph.D study in the UK for the period 2003-2007. He has recently completed his Ph.D thesis “Trade liberalization and labour market outcomes in Vietnam” at the University of Sussex. Hung has worked intensively with the VLSSs and VHLSSs and most of the econometric models proposed in this research. His research interests focus on trade and poverty in Vietnam, on which he has had some working/conference papers, and three papers are currently under revision for *Journal of Comparative Economics, Journal of Asian Economics,* and *Review of Income and Wealth* (see CV).

Tuan is one of the few ‘old generation’ economists at the NEU who were not trained by orthodox Marxist economics (he got his first postgraduate degree in the Republic of Korea). With regard to the focus of this study, Tuan has worked extensively on the impact of foreign direct investment on job creation in Vietnam (see CV). Beside his current position as the Dean of the Faculty of International Education, Tuan is also an associate professor at the Faculty of Labour Economics, where he holds the professor chair of organizational behaviour. Given this, Tuan has a wide contact with officials at the MOLISA, MPI, and MARD. In addition, he has also worked with some leading donors in Vietnam such as Sida and ADB. The participation by Tuan in this project will have two important contributions. First, with his expertise in labour economics, Tuan will provide a ‘critical eye’ on the analysis of the research reports. Second,
Tuan will be mainly responsible for the ‘policy dialogue’ component of this project. In this regard, Dr. Tuan will organize and coordinate the ‘policy dialogue’ as explained above.

Thanh is a young lecturer and researcher at NEU. Thanh has got her master in University of Toulouse 1 (France), with a strong background in microeconomic analysis, quantitative methods, and competence in using statistic softwares such as STATA and SPSS. Thanh has been recruited as a lecturer at NEU’s Faculty of Economics and a researcher at Faculty of International Education since January 2004. Since then, she has acquired certain research experience in the issues related to the agricultural sector in Vietnam at the threshold of the WTO accession (see CV). Her participation in this project will be a first-hand experience in undertaking empirical analysis using the household surveys. In this regard, her skills in quantitative analysis learned during her study in Toulouse would be a useful asset.

9. Expected capacity building

As the research team exhibits a mix of seniority (Tuan) and youth (Hung and Thanh), the expected capacity building described below is mainly targeted for Hung and Thanh.

Hung and Thanh are young lecturers at NEU, a leading university in Vietnam in the areas of economics and management. In common with other research and training institutions in Vietnam, NEU has been in a special transition from ‘Marxist’ economists, who were trained by the orthodox Marxist doctrine in universities in Beijing, Moscow, former East Germany, etc. (with a few exceptions including Tuan) and the ‘new generation’ economists, who have recently acquired ‘capitalist’ economics in the US, Canada, the UK, Australia. The systematic differences between the two generations make it difficult for the young economists like Hung and Thanh to rely on most of their senior colleagues for knowledge transfer and capacity building. This is reflected in the NEU’s recent strategy to build up their human resource (NEU, 2005). This strategy emphasizes the university’s policies to seek for technical and financial assistance from international donors to “[…] create learning-by-doing opportunities for NEU junior staff by working with international experts though a number of research activities assisted by international donors” (p.12 – translated from Vietnamese).

Though this research proposal is the researchers’ own effort, this accords to the NEU’s strategy and is expected to contribute to it. For Hung as the lead researcher this study is complementary to his just-finished Ph.D thesis at the University of Sussex. The study will explore the same datasets to one of the issues listed in the conclusion part of his Ph.D thesis as future research schedule. In this regard, this research is a part of his continuing ‘learning-by-doing’ process. Given his experience in working on the trade reform, rural poverty, and labour market issues in
Vietnam, this study will enhance his research expertise in these areas, and especially to build up a publication track record on the research area.

For the team members, Tuan is willing to participate in this project to mainly provide a ‘critical eye’ and contacts necessary for the ‘policy dialogue’ component. His in-depth knowledge in the areas will provide a good source of reference for the lead researcher and other team member. This study will provide a chance for Thanh to get her first-hand experience in analyzing the household surveys. In this regard, she will learn from the experience of the lead researcher in undertaking ‘data crouching’, which is probably important to consolidate a necessary skill in applied research using econometric analysis.

Since this is a small project involving one country and three researchers, a sophisticated project management plan is probably not necessary. Since the research team has been working with each other in other capacities (see their CVs), the interaction between the research team in this study means an unusual knowledge transfer – which is the informal and natural knowledge transfer between closely-teamwork colleagues. We believe that this unusual and friendly channel of knowledge transfer will work in this project.

10. Ethical, social, gender or environmental issues or risks

This study does not incur any ethical, social, gender or environmental issues or risk.

11. List of past, current or pending projects in related areas involving team members

- ‘Trade Liberalization and Industrial Pollution in Vietnam’, funded by IDRC (though Economy and Environment Program for Southeast Asia - EEPSEA), from Oct 2006 to Jan 2008. Grant No.: 003591-104. Team members: Thai Hung Pham (lead researcher), Nguyen The Chinh (researcher, NEU), Bui Anh Tuan (researcher, NEU).

- ‘Ethnicity Underdevelopment in Vietnam’, funded by DfID (Department of International Development, UK) and ESRC (Economic and Social Research Council, UK), from Dec 2006 to Dec 2007. Team members: Bob Baulch (project leader, IDS), Barry Reilly (researcher, Sussex), Thai Hung Pham (researcher).
Appendix 1:
Nonfarm Diversification and Income in Rural Vietnam

Table A1: Incidence of Nonfarm Diversification in Rural Vietnam, 1993-2002

<table>
<thead>
<tr>
<th>Nonfarm diversification at the individual level (%)*</th>
<th>1993</th>
<th>1998</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>77.4</td>
<td>71.2</td>
<td>64.4</td>
</tr>
<tr>
<td>Nonfarm wage employment</td>
<td>11.1</td>
<td>14.7</td>
<td>20.1</td>
</tr>
<tr>
<td>Nonfarm self employment</td>
<td>11.4</td>
<td>14.1</td>
<td>15.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonfarm diversification at the household level (%)**</th>
<th>1993</th>
<th>1998</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm-only household</td>
<td>64.7</td>
<td>55.2</td>
<td>45.8</td>
</tr>
<tr>
<td>Diversified household</td>
<td>23.8</td>
<td>29.5</td>
<td>34.7</td>
</tr>
<tr>
<td>Nonfarm-only household</td>
<td>11.4</td>
<td>15.3</td>
<td>19.5</td>
</tr>
</tbody>
</table>

Source: calculations from the VLSS 1992/93, VLSS 1997/98, and VHLSS 2002

Notes:

a. * Individual employment outcomes are specified on the basis of the primary (most time-consuming) jobs. ‘Agriculture’ refers to those who cultivate in their farms or are hired by the others to work on their farms as traditional agricultural activities; ‘Nonfarm wage employment’ includes those who are wage-employed in nonfarm activities in the rural areas; and ‘nonfarm self employment’ implies those who are self-employed in nonfarm activities.

b. ** Household nonfarm diversification outcomes are defined on the basis of the employment statuses of their household members (as above). ‘Farm-only’ households, or completely undiversified, are those with all household members working in agriculture as their main occupations. The nonfarm-only households are completely diversified into the RNFS and hence none of their members work in agriculture. In between these two categories, the diversified households allocate their labour supply for both farm labour and nonfarm activities.

c. It should be noted that the employment outcomes listed in this table are broadly defined mainly to highlight the relative importance of nonfarm activities. Further discussions are subject to the latter stage of this research.
### Table A2: (Preliminary) Shares of Rural Household Income Sources in Vietnam, 1993-2002 (%)

<table>
<thead>
<tr>
<th></th>
<th>1993</th>
<th>1998</th>
<th>2002</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rural sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture income</td>
<td>50.36</td>
<td>53.89</td>
<td>32.99</td>
<td>45.75</td>
</tr>
<tr>
<td>Nonfarm wage income</td>
<td>11.35</td>
<td>10.85</td>
<td>19.12</td>
<td>13.77</td>
</tr>
<tr>
<td>Nonfarm self-employment income</td>
<td>25.08</td>
<td>20.87</td>
<td>27.35</td>
<td>24.43</td>
</tr>
<tr>
<td>Other income sources</td>
<td>13.21</td>
<td>14.39</td>
<td>20.53</td>
<td>16.05</td>
</tr>
<tr>
<td><strong>Lowest quintile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture income</td>
<td>58.74</td>
<td>66.80</td>
<td>39.99</td>
<td>55.18</td>
</tr>
<tr>
<td>Nonfarm wage income</td>
<td>13.70</td>
<td>10.35</td>
<td>15.49</td>
<td>13.18</td>
</tr>
<tr>
<td>Nonfarm self-employment income</td>
<td>19.05</td>
<td>12.97</td>
<td>29.70</td>
<td>20.57</td>
</tr>
<tr>
<td>Other income sources</td>
<td>8.51</td>
<td>9.88</td>
<td>14.81</td>
<td>11.07</td>
</tr>
<tr>
<td><strong>Middle quintile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture income</td>
<td>56.03</td>
<td>53.70</td>
<td>34.09</td>
<td>47.94</td>
</tr>
<tr>
<td>Nonfarm wage income</td>
<td>10.15</td>
<td>10.95</td>
<td>20.02</td>
<td>13.70</td>
</tr>
<tr>
<td>Other income sources</td>
<td>12.70</td>
<td>13.60</td>
<td>19.25</td>
<td>15.18</td>
</tr>
<tr>
<td><strong>Highest quintile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture income</td>
<td>32.66</td>
<td>38.02</td>
<td>23.83</td>
<td>31.50</td>
</tr>
<tr>
<td>Nonfarm wage income</td>
<td>12.77</td>
<td>11.47</td>
<td>20.44</td>
<td>14.89</td>
</tr>
<tr>
<td>Nonfarm self-employment income</td>
<td>36.10</td>
<td>28.38</td>
<td>26.75</td>
<td>30.41</td>
</tr>
<tr>
<td>Other income sources</td>
<td>18.47</td>
<td>22.12</td>
<td>28.99</td>
<td>23.19</td>
</tr>
</tbody>
</table>


**Notes:**

a. These sources of incomes are defined as follows:
   - Agriculture incomes include net income (i.e., total production value minus expenditures) from crops, livestocks, forestry, and aquaculture.
   - Nonfarm wage income is defined as all sorts of payments, including salaries, allowances, bonus, in cash and in kind that household members have received.
   - Nonfarm self-employment income consist of incomes from nonfarm household business; processing of household produced crops at home; incomes from providing farm-related services; rents received from letting lands, tools, animals…
   - Other income sources cover any scholarships/awards, remittances, pension payments, subsidies, remittances, selling property, lottery, withdrawing from savings, and any other non-labour incomes.

b. Computing agriculture incomes (and some nonfarm products) requires converting the crop output given in quantity into monetary terms. Due to the back of price data, ‘unit values’ defined consistently with Deaton (1988) are used convert the output data from quantities into monetary values. The unit values are common alternatives when data on prices is either noisy or not sufficient (which is actually the case in the Vietnamese household surveys) (see Niimi et al. (2003), Litchfield et al., (2006) for more details).

c. Calculating income data from household living standard surveys in developing countries is widely recognized as a complicated procedure with relatively low level of accuracy due to several factors. Apart from the obvious reason that respondents interviewed in these surveys generally do not provide precise estimates of their incomes, the fact that rural households rely on a diversified portfolio of income-generating activities make this process even more complicated. The calculation procedure in this case based on the textual description of each types of incomes (as provided in the questionnaires and other related documents) to ensure these derived figures as precise as possible. However, it should be noted that these are preliminary results and thus interpreted with great cautions (STATA do files on this calculation procedure are available from the authors upon request).
Table A3: Share of Nonfarm Income in Vietnam and Other Developing Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Years</th>
<th>Share of nonfarm incomes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Africa (average)</strong> a</td>
<td>(various)</td>
<td>42</td>
</tr>
<tr>
<td>Botswana</td>
<td>1985-86</td>
<td>77</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>1981-84</td>
<td>37</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1989-90</td>
<td>36</td>
</tr>
<tr>
<td>Kenya (central)</td>
<td>1974-75</td>
<td>42</td>
</tr>
<tr>
<td>Tanzania</td>
<td>1980</td>
<td>25</td>
</tr>
<tr>
<td>Uganda (Mbale district)</td>
<td>2001</td>
<td>50</td>
</tr>
<tr>
<td><strong>Latin America (average)</strong> b</td>
<td>(various)</td>
<td>40</td>
</tr>
<tr>
<td>Chile</td>
<td>1990</td>
<td>32</td>
</tr>
<tr>
<td>Mexico</td>
<td>1992</td>
<td>50</td>
</tr>
<tr>
<td>Mexico</td>
<td>2002</td>
<td>76</td>
</tr>
<tr>
<td>Ecuador</td>
<td>1995</td>
<td>41</td>
</tr>
<tr>
<td><strong>Asia (average)</strong> a</td>
<td>(various)</td>
<td>32</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1988-89</td>
<td>31</td>
</tr>
<tr>
<td>India</td>
<td>1993-94</td>
<td>34</td>
</tr>
<tr>
<td>China (Guangdong)</td>
<td>1989</td>
<td>34</td>
</tr>
<tr>
<td>Philippines (Mindanao)</td>
<td>1984-85</td>
<td>23</td>
</tr>
<tr>
<td><strong>Vietnam</strong></td>
<td>1993-2002</td>
<td>38</td>
</tr>
<tr>
<td>Vietnam</td>
<td>1993</td>
<td>36</td>
</tr>
<tr>
<td>Vietnam</td>
<td>2002</td>
<td>46</td>
</tr>
</tbody>
</table>


Notes:

a. As these figures were reported using different definitions of nonfarm income sources from the surveys with distinctive scales and techniques, they are thus subject to differences in measurement method and should be interpreted with caution.

b. For brevity, not all of the above sources are cited in the short list of key references in this proposal. Full references with details of these papers are available from the authors upon request.
Notes: see notes in Table A1 for the definitions of these household groups.
Figure A2: (Preliminary) Real Nonfarm Income Per Capita, 1993-2002


Notes: ‘nonfarm income’ in this figure refers to wage income and nonfarm self-employment incomes (see notes in Table A2 for the definitions of these household groups).
Appendix 2: Technical Appendix

This appendix outlines the econometric methods used to investigate the project’s research questions (question 1 to 3) using the Vietnam Living Standards Surveys (VLSS) 1992/93; 1997/98, and the Vietnam Household Living Standards Surveys (VHLSS) 2002 and 2004. (see sub-section Policy dialogue in the main text for question 4).

**Question 1: How important is nonfarm income for rural households? To what extent nonfarm income has contributed to overall rural poverty and income inequality?**

The Shapley approach is employed to examine to what extent nonfarm income has contributed to overall rural poverty and income inequality in Vietnam (see Duclos and Araar, 2006 for an arguably the most comprehensive discussion of the Shapley approach in poverty and inequality analysis). This framework can be outlined based on Duclos and Araar (2006), and Araar (2006) as follows.

Consider \( n \) players in a given set \( N \) in the theory of cooperative games, they might form coalitions (these are subset \( S \) of \( N \) in the games to obtain the surplus and then redistribute it among themselves. Assume that such coalitions are determined by a function \( v \), the main question to resolve is how the surplus can be divided between \( n \) players given \( v \). According to the Shapley approach, the value or the expected gain of player \( k \), noted by \( E_k \) is then given by the following formula:

\[
E_k = \sum_{S=S}^{s=\{0, n-1\}} \frac{s!(n-s-1)!}{n!} MV(S, k) \quad \text{with} \quad MV(s, k) = [v(s \cup \{k\}) - v(s)]
\]

(1)

where the term \( MV(S, k) \) is the marginal value that the play \( k \) generates after his adhesion to the coalition \( S \). As there are many different possibilities of coalitions, the next question is what will be the expected marginal contribution of player \( k \)? Note that the size of the coalition \( S \) is limited to \( s \in (0, 1, ..., n-1) \) and suppose that the \( n \) players are randomly ordered such that

\[
\sigma = \left( \sigma_1, \sigma_2, ..., \sigma_{i-1}, \sigma_i', \sigma_{i+1}', ..., \sigma_n' \right)_{n-s-1} \quad \text{with} \quad \sigma \text{ denotes the order}
\]

After some simulation, Araar (2006) shows that the expected marginal contribution of player \( k \), given by the Shapley value, can be expressed as:

\[
E_k = \frac{1}{n!} \sum_{i=1}^{n!} MV(\sigma', k)
\]

(2)

where for each order \( \sigma \) of the \( n! \) orders, the player \( k \) has only one position that determines the coalition to which he or she can adhere; and the term \( MV(\sigma', k) \) represents the marginal value that player \( k \) add to the surplus when he or she is adhered to that coalition.
Although the framework was originally developed in the theory of cooperative games, this can be applied in a number of contexts (see Shorrocks, 1999; Duclos and Araar, 2006). In this study, the Shapley approach will be used to decompose the poverty and inequality indices by income sources. By supposing that income sources represent factors that contributes to the poverty (the FGT(n) index for instance) or income inequality (such as the Gini) index (or coefficient), the expected marginal contribution of income source \( g \), according to the Shapley approach, is equal to the following:

\[
E_g^G = \frac{1}{G!} \sum_{i=1}^{G} MV(\sigma^i, g)
\]

where \( \sigma^i \) represents the \( i^{th} \) possible order of sources and \( MV(\sigma^i, g) \) represents the impact of eliminating source \( g \) for the order \( \sigma^i \) on the contribution of the set of source \( G \). Such decomposition can then be easily implemented using the DAD software (see section 13 in the User Manual for DAD 4.4 by Duclos and Araar, 2004 for more details).

**Question 2: What are determinants of nonfarm diversification by rural households?**

**Multinomial logit (MNL) models**

Suppose that a rural household might be attached to one of the three different employment outcomes, including (i) solely working in agriculture; (ii) working in both agriculture and the RNFS; and (iii) solely working in the RNFS. These three exclusive outcomes are defined on the basis of primary employment by household members. Let’s \( Y = j \) if the \( i^{th} \) household is in the \( j^{th} \) alternative employment outcome, the probability that household \( i \) experiences (unordered) outcome \( j \) is expressed as (subscript \( i \) is suppressed for simplicity):

\[
P(Y = j) = \frac{e^{X\beta_j}}{\sum_{j=1}^{3} e^{X\beta_j}} \quad \text{for } j = 1, 2, 3
\]

where \( P(Y=j) \) with \( j = 1, 2, 3 \) represents the probability of a household being in either one of the three diversification outcome above; \( X \) is a \((k\times1)\) vector of characteristics for each household in the sample; \( \beta_j \) is a \((k\times1)\) vector of coefficients on \( X \) applicable in state \( j \).

The vector \( X \) includes a number of household-level, and community-level characteristics. Though the detailed construction of the vector \( X \) is not described here, as suggested by the literature, we will construct the vector \( X \) to reflect family background (parental education, religions, and ethnicity); demographical characteristics (household structures, fertility ratios); household physical assets (pre-determined ones); household landholdings; household access to credit, non-labour income sources; and community-level characteristics (physical and institutional infrastructures, general economic conditions, access to supports from authorities, and other agro-climate characteristics).
The Theil normalization will be applied to the first outcome (solely working in agriculture). The Small-Hsiao test, as discussed in Wills (1987), will be used to test the IIA property, and the Wald test will be adopted to test whether any two outcomes out of the three unordered employment outcomes under consideration can be combined. In the case the MNL model fails to pass the above specification test of the IIA property, the nested logit (NL) model will be applied as alternative. The NL models are currently preferred to the simple MNL model (see Hensher and Greene, 2002 for a review). Although the NL model allows interdependence between pairs of alternatives, this model is more complicated to estimate and the results depend on the assumption made with respect to the utility function. Koppleman and Wen (1998) and Siberhorn et al. (2006) demonstrate that the two unequal forms of the NL models, which are the utility maximization nested logit (UMNL) and non-normalized nested logit (NNNL) model have different properties which impact the estimation results. In this case, the four above outcomes will be nested in various ways to estimate both the UMNL and NNNL models. The details of these NL models are not outlined here for brevity. Full exposition of these models can be found in, for instance, Koppleman and Wen (1998), Hensher and Greene (2002), Siberhorn et al. (2006).

Regarding the application of the Krueger and Summers’s (1988) method, variations of the probabilities by regions from the mean (instead of comparing with an arbitrarily omitted region as the base category) will be retrieved from estimating the MNL reduced form models or the NL models (with certain modification as suggested in Haisken-DeNew and Schmidt, 1997). The robust standard deviations of these coefficients will be estimated by the method suggested in Zanchi (1998) (see Pham 2007, Pham et al., 2007 for the application of this method in the Vietnamese context).

Quantile regression models

In the MNL models, nonfarm diversification by rural households is necessarily measured as a polychotomous indicator. It is also possible to model the incidence of diversification as the share of household members employed in the RNFS. However, as this variable is only observed for the households who actually participate in the RNFS, our samples are thus truncated at zero. If the process of participation in the RNFS sector is not random, ignoring this truncation will result in biased estimates. Therefore, it is important to correct for selection bias. In this research, we will use the Heckman’s (1979) two-stage procedure to overcome this problem.

In the first stage, a probit model of whether a household is diversified into the RNFS is estimated on a set of the characteristics included in vector \(X\) (see equation 4) to obtain the inverse of the Mill’s ratio, \(\lambda\).

In the second stage, the following equation, corrected for selection bias, will be estimated using OLS:

\[
\frac{\partial \ln y_{i}}{\partial x} = \beta' + \gamma' \lambda
\]

It should be noted that the application of the NL model in this context is subject to further consideration. We appreciate the suggestion of the reviewer(s) on the earlier version of this proposal on the possibility of applying the NL model, instead of the simple MNL one. To the authors’ knowledge, the NL model is predominantly used in the areas of transportation, logistics, and marketing (such as brand choice, purchase decisions) (see Silberhorn et al. 2006 for a review). In principles, the NL model can be used in any situation where subsets of alternatives share unobservable utility components. The NL model will then classifies alternatives into nests on the basis of their similarities and the main goal of this process is to accommodate the violation of the IIA assumption. However, given the alternative outcomes specified in (4) and (5), it is not straightforward to justify the nesting among these alternatives (i.e. which alternatives should be nested and what are the rationale underlying such nesting).
Where \( Y \) is the share of household members employed in the RNFS, \( H \) is a vector of household characteristics, where \( \ddot{\lambda} = \rho_j \sigma_{u_j} \) is the coefficient of the selection bias correction term; \( \rho_j \) is the coefficient of correlation between the error terms in the income equation and the probit selection equation (estimated in the first stage). The construction of vector \( H \) is subject to further discussion at a latter stage of this research. In general, this vector will be specified essentially the same as Baulch, Pham, and Reilly (2007).

An exclusive focus on the determinants of nonfarm diversification estimated at the average as in (5) may provide misleading impacts of covariates across the conditional distribution. With a quantile regression approach, the focus moves away from the mean to selected points on the conditional distribution and the estimation procedure is formulated in terms of minimizing absolute rather than squared errors, and the estimator is known as the Least Absolute Deviations estimator. The quantile regression approach is less sensitive to outliers and provides more robust estimators in the face of departures from normality than the OLS technique (see Koenker, 2005; Koenker and Bassett, 1978). According to Deaton (1997) quantile regression models may also have better properties than OLS in the presence of heteroscedasticity.

In this research, the quantile regression approach is used to take into account the heterogeneity of nonfarm diversification. By investigating the impacts on nonfarm diversification at different points of the conditional distribution of diversification, we hope to provide insights on how households diversify differently into the RNFS. The quantile regression approach applied for the equation (5) at \( \theta \)th quantile, using the terminology of Koenker and Hallock (2001), can be written as:

\[
Y = H' \ddot{\lambda} + h_\theta(v) + \mu_\theta \quad (0 \leq \theta \leq 1) \tag{6}
\]

where \( h_\theta(v) \) is the Buchinsky’s (1998) polynomial series expansion. In pursuing this approximation, the constant term of the wage equation will be conflated with the non-zero constant term of the polynomial series expansion, and therefore these two constant terms cannot be identified separately. Buchinsky (1998) provides a solution to this identification problem by using a subset of observations for which the participation probability is close to one to identify the constant term of the polynomial series expansion. However, this method effectively requires a relatively large sample (Buchinsky (1998) using the samples that range from 39,746 to 46,866 observations), but this is constrained by the data availability in the current study. Given this, this research employs a crude expedient of inserting the Heckman’s selection bias correction term (as above) and its quadratic term into the quantile regression models. It is acknowledged that, in contrast to the mean regression case, this provides an inexact correction for selection bias. However, it circumvents the much trickier problem of identifying the quantile wage regression constant terms.
Question 3: What are the determinants of nonfarm income? To what extent does nonfarm diversification contribute to improve rural household welfare? And does nonfarm diversification benefit differently for the poor and for the non-poor?

Quantile regression models

The use of quantile regression approach in this context is to examine the determinants of nonfarm income at different points of the conditional distribution of nonfarm income. This is expected to provide evidence on how the income covariates affect differently lower-income households located at the bottom ends and higher-income households located at the top ends of the conditional distribution of income. The specification of the quantile regression model with correction for selection bias is similar to equation (6). Instead of having the incidence of nonfarm diversification, this section will examine real per capita nonfarm income. As ‘other income sources’ (see Table A2) are non-labour sources of incomes, we excluded this category from nonfarm income examined in this section. In Figure A2 of the Appendix 1, we observed the pattern of distribution of nonfarm income per capita for the three years from 1993 to 2002.

Two-stage Least Squares Approach

The most general structural form of the welfare function of household $i$ is as follows

$$ Y_i = \beta_0 + E_i'\beta_1 + X_i'\beta_2 + u_i $$

(7)

where $Y_i$ is the consumption expenditure level of the household $i$; $E_i$ is the share of household members employed in the RNFS or the share of nonfarm income; $X_i$ is a column vector defined as in equation (4).

The potential endogeneity of some variables in the $X_i$ vector can be resolved by using the initial period characteristics to define this vector. In this case, the structural equation (6) can be reduced to the following equation:

$$ \Delta Y_i = \alpha_0 + \Delta E_i'\alpha_1 + X_{i,t-1}'\alpha_2 + \mu_i $$

(7)

where $X_{t-1,i}$ is the $X_i$ vector defined by the characteristics of the initial period ($t-1$); $\Delta Y_i$ is the change in household expenditures; and $\Delta E_i$ is the either change in the share of the household members employed in the RNFS between the two years or the change in the share of nonfarm income (expressed in fractional points). The rest of this sub-section described the method using $\Delta E_i$ is the change in the share of household members employed in the RNFS between the two years. It is also equally applicable when $\Delta E_i$ is the change in the share of nonfarm income.

---

11 This study will use both the share of household members employed in the RNFS and the share of nonfarm income as a measure for nonfarm diversification in the TSLS approach. For brevity, the discussion here focuses on the share of household members employed in the RNFS, but it is equally applicable to the share of nonfarm income.
However, the potential endogeneity of $\Delta E_i$ in equation (7) needs to be resolved. The TSLS is then applied to replace the actual problematic $\Delta E_i$ variable in that equation by a counterpart variable that is purged of its stochastic or random component to ensure that the OLS procedure could be applied. In order to do this a ‘reduced form’ equation is specified, in which the change in the share of household members employed in the RNFS is specified as a function of all the exogenous variables in equation (7) (i.e. the vector $X_{i,t-1}$) and a set of instrumental variables as:

$$
\Delta E_i = \delta_0 + X_{i,t-1}'\delta_1 + Z_i'\delta_2 + \epsilon_i
$$

where $Z_i$ is a vector of instrumental variables, which exert impacts on nonfarm diversification but not on household expenditures. The predicted values from this OLS-estimated ‘reduced form’ equation (8), defined as $\hat{\Delta E}_i$, is then inserted in the structural equation to replace the problematic $\Delta E_i$. As a result, equation (7) can be reduced to the following reduced-form equation that can be then estimated by using OLS:

$$
\Delta Y_i = \partial_0 + \hat{\Delta E}_i'\partial_1 + X_{i,t-1}'\partial_2 + \sigma_i
$$

This study will use a number of instruments that reflect the change in the availability of nonfarm opportunities and the change in the demand side of nonfarm labour at the commune level between the two years of the panel under consideration. To verify the validity of instruments, an F-test for instrument relevance (using Stock and Watson’s (2003) ‘rule-of-thumb’), the Sargan’s test for exogeneity of instruments, and the Hausman specification test for the exogeneity of the problematic regressor are necessary.

**Propensity Score Matching Approach**

Let’s $Y_{i1}$ presents the value of the expenditure level of the $i^{th}$ diversified household (also called treatment) and $Y_{i0}$ represents the expenditure level would be if this household did not participate in the RNFS (also called control). Let $D_i$ be a treatment indicator, which is equal to one if a household is diversified and zero otherwise. Then the observed welfare outcome of the $i^{th}$ household is $Y_i = D_iY_{i1} + (1 - D_i)Y_{i0}$ and the welfare effect of being diversified for a single household is $\tau_i = Y_{i1} - Y_{i0}$.

In a non-experiment study, the selection of households in the rural population into different employment outcomes is not random. In this context, the $i^{th}$ household can only either diversifies or not into the RNFS, therefore only one of $Y_{i1}$ or $Y_{i0}$ can be actually observed. This problem of unobservability implies that while $E(Y_{i1}|D_i = 1)$ can be estimated, it is not possible to estimate $E(Y_{i0}|D_i = 1)$. In other words, the

---

12 To simplify the framework, in this section, the treatment group is defined by the households who diversified into the RNFS and the control group comprises of those who do not participate in the rural nonfarm labour market. When applying this PSM approach to examine the welfare effect of nonfarm diversification, the definitions of these treatment and control groups will be changed.
welfare effect of nonfarm diversification, now measured as the average treatment effect (\(ATE\)), is given in the non-experiment setting as

\[
\tau\big|_{D_i=1} = E(Y_{it} | D_i = 1) - E(Y_{it} | D_i = 0)
\]

(10)
cannot be estimated.\(^{13}\)

Rubin (1977) makes a proposition that extends the experimental framework to non-experimental studies. This proposition states that if for each household we observe a vector of pre-treatment covariates \(T_i\), the assignment to the treatment is then assumed to be associated only with this pre-treatment vector. This assumption, which is also called the ignorable treatment assignment assumption or selection on observables, implies that conditional to the vector \(T_i\), \(Y_i\) or \(Y_{it}\) are orthogonal to the treatment indicator \(D_i\) with all \(i\). In other words,

\[
E\left(Y_i \big| T_i, D_i = 1\right) = E\left(Y_i \big| T_i, D_i = 0\right) \quad \text{with } j = 0, 1
\]

which means that conditional to the observable \(T_i\) there is no systematic pre-treatment difference between the diversified group and the undiversified group. Under this assumption, the \(ATE\) can be expressed as:

\[
\tau\big|_{D_i=1} = E\left\{E(Y_{it} \big| T_i, D_i = 1) - E(Y_{it} \big| T_i, D_i = 0) \big| D_i = 1\right\}
\]

(11)
where the outer expectation is over the distribution of \(T_i | D_i = 1\), the pre-treatment variables in the treated population.

The welfare effect of nonfarm diversification can be then estimated through matching the diversified households with undiversified counterparts on the vector \(T_i\). Effectively, it implies assigning observations into cells defined by unique values of the covariates. However, this matching is difficult when the vector \(T_i\) is highly dimensional, and especially when it includes some continuous variables. In this context, Rosenbaum and Rubin (1983) develop another proposition which allows the use of the propensity score, defined as the conditional probability of receiving a treatment (i.e. being diversified in this case) given a set of covariates, to reduce the dimensionality of this matching. Let’s \(p(T_i)\) is the conditional probability of the \(i^{th}\) household being diversified in the RNFS given the pre-treatment vector \(T_i\). This propensity score of \(i^{th}\) household to diversify is defined as

\[
p(T_i) = \Pr(D_i = 1 | T_i) = E(D_i | T_i) \quad \text{with } 0 < p(X_i) < 1 \text{ for all } T_i, \text{ and}
\]

\(^{13}\) Obviously, the difference \(\tau^e = E(Y_{it} | D_i = 1) - E(Y_{it} | D_i = 0)\) can be estimated. This would only be a solution in the experimental setting, where the treatment assignment is drawn randomly from the same population. In this case, there would be no systematic pre-treatment difference between the diversified and undiversified groups, making the conditioning on \(D_i\) unnecessary. However, in a non-experimental study, this estimation is then a biased estimator of the \(ATE\).
\[ \Pr(D_1, D_2, \ldots, D_N | T_1, T_2, \ldots, T_N) = \prod_{i=1,2,\ldots,N} p(T_i)^{D_i}(1 - p(T_i))^{(1-D_i)}. \]

Rosenbaum and Rubin (1983) then argue that if there are no systematic pre-treatment differences between the treatment and the control groups, conditional to the observable \( T_i \), there would also be no systematic pre-treatment differences between the treatment and the control, conditional to the propensity score, \( p(T_i) \). In other words, conditional to the \( p(T_i) \), \( Y_i \) or \( Y_0 \) are orthogonal to the treatment indicator \( D_i \) with all \( i \).

Combining the Rubin’s (1977) proposition (in expression 11) and this Rosenbaum and Rubin’s (1983) proposition, the \( ATE \) is now given as:

\[ \tau_{\bar{D}_i=1} = E[Y_i | p(T_i), D_i = 1] - E[Y_i | p(T_i), D_i = 0, D_i = 1] \]  \hspace{1cm} (12)

assuming that the expectations are defined, the outer expectation is over the distribution of the propensity score in the treated population, \( p(T_i) | D_i = 1 \). Moving from (11) to (12) means a move from estimating the \( ATE \) conditional to the pre-treatment observable characteristics, \( T_i \), to deriving that \( ATE \) conditional to a single propensity core index, \( p(T_i) \).

Given this PSM method, the estimation strategy is carried out in the two steps. The first step involves estimating the propensity score by using either a logit or a probit model, using the \( T_i \) vector as the set of explanatory. The second step involves matching the treatment (or the diversified households) with the control (or the undiversified households) based on the estimated propensity scores. Several non-parametric matching techniques have been proposed, including Nearest Neighbour Matching, Radius Matching, Kernel Matching, and Stratification Matching as explained in Becker and Ichino (2002); and Local Linear Regression Matching, Spline Matching, and Mahalanobis Matching as shown in Leuven and Sianesi (2003).

Does nonfarm diversification benefit differently for the poor and for the non-poor? To examine this question, the TSLS and PSM approaches will be carried out for the poor sub-sample and the non-poor sub-sample (after testing if such separation is statistically justifiable). The identification of the poor and the non-poor will be made using (i) the official poverty line; (ii) 10% below (and above) poverty lines; and (iii) using adult equivalent units to adjust for possible differences in consumption behaviours among the household members. These adjustments will also allow for a test of sensitivity of the estimation results using the TSLS and the PSM approaches.