# Essays on horizontal inequalities, ethnicity and poverty : The case of Vietnam

Thuy Quynh Hoang

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

**Préparée à l'Université Paris-Dauphine**

# Essays on horizontal inequalities, ethnicity and poverty: The case of Vietnam

**École Doctorale de Dauphine — ED 543**

Spécialité **Sciences économiques**

**COMPOSITION DU JURY :**

M. François ROUBAUD
IRD-DIAL, Université Paris -Dauphine
Co-Directeur de thèse

M. Alain DESDOIGTS
IEDES, Université Paris 1
Rapporteur

Mme. Dominique van de WALLE
Banque mondiale
Rapporteur

M. Michael GRIMM
Université de Passau
Président du jury

Mme. Marta MENENDEZ
Université Paris -Dauphine
Membre du jury

**Soutenue le 30.11.2018**
**par Quynh HOANG**

Dirigée par **Mireille RAZAFINDRAKOT**
**François ROUBAUD**

DAUPHINE UNIVERSITÉ PARIS | PSL

Paris Sciences & Lettres

# PH.D THESIS

Submitted for the degree of
Doctor of Philosophy of Paris Sciences & Lettres in Economics

Prepared at Université Paris Dauphine and IRD-DIAL and defended at Université Paris Dauphine on Nov. 30, 2018

by

Quynh HOANG

# Essays on horizontal inequalities, ethnicity and poverty: The case of Vietnam

Thesis Advisors: Mireille RAZAFINDRAKOTO & François ROUBAUD

Jury:

| | | |
|---|---|---|
| *Reviewers:* | Alain DESDOIGTS | IEDES, Pantheon-Sorbonne University |
| | Dominique van de WALLE | The World Bank |
| *Advisor:* | François ROUBAUD | IRD-DIAL, Paris-Dauphine University |
| *Examiners:* | Michael GRIMM | University of Passau |
| | Marta MENENDEZ | Paris-Dauphine University |

*To my beloved grandparents*

# Acknowledgments

I am deeply grateful to my supervisors, Mireille Razafindrakoto and François Roubaud, for their incredibly helpful guidance and endless support throughout my PhD. They first encouraged me to apply for the PhD program, then, have inspired me and kept their steadfast confidence in me on this intellectual journey even when I doubt myself. I would like to express my gratitude to Dominique van de Walle and Alain Desdoigts for accepting to be my referees and for their extremely useful comments during my pre-defense. My sincere thanks also go to Michael Grimm and Marta Menendez for being part of my thesis Jury.

I am also indebted to my co-authors, Laure Pasquier-Doumer and Camille Saint-Macary, for giving me the chance to work with them, to learn from them and for their direct contributions in my work. Their selfless time, feedback and support are beyond my expectation. I thank Axel Demenet also for his friendship and unwavering support as my "đàn anh"/PhD senior and my co-author in another work (not included in this thesis).

My sincere thanks go to all of my DIAL team; their support, each one in their own way, has given me such a good time. It is Danielle who is so patient to speak in French with me. It is Anne who always welcomes me every morning with a nice smile and gave me many cookies when I shared the office with her. It is Loic who has helped me a lot for administrative procedures, particularly in managing my conference trips. I thank Flore, Sandrine, Lisa, Christophe, Véronique, Anne Sophie, Jean-Michel, Jean-Noël, Dali, Helene, Charlotte, Anda, Claire, Delia, Sarah, Linda, Niri, Luis, Marin, Björn, Quentin, Estelle, Esther, Raphael, Virginie, Thomas (thanks for giving me chocolate during my last week of thesis submission), Siyavash (your support for my last night at DIAL working on the thesis was precious), Ana, Tijan, Thomas T., Rawane, Tam, Dat (thanks for your generous sharings) and Phuong. I have received encouragements from all of DIALiens through coffee breaks, lunch time, Friday drinks, football matches (inside and outside the office!), World cup watching afternoons, Christmas parties, and movie and karaoke nights. My special thanks go to Anne, Oscar and Marlon, who are not only my DIAL colleagues, my conference buddies, my thesis editors but also my dearest friends.

I also acknowledge the friendship and support of my friends outside the office but not less important. Thay Cuong Le Van, thanks for being supportive and giving me many opportunities to present my work in Vietnam. Xavier and Danielle, thank for your amazing kindness and warmest support, for receiving me at your home in Hanoi, Paris and Brittany. Linh, thanks for being my first friend in Paris and helping me endlessly from the master to the PhD. I also thank Olivier Ruchet and Vincent Fertey for giving me the opportunities to teach at Sciences Po and Charles Oman for accepting me as his teaching assistant. I appreciate the incredible support from my former boss and my colleagues at the Development Research and Consultancy Centre in Hanoi, especially in helping me doing field work in 2015. I thank Johann for his sweet, supportive, enthusiastic and artistic presence in the last few months of my PhD. I am also grateful to all of my friends (in Vietnam, France and the US) who have dealt with my troubles, my complaints, my doubts but still encouraged me strongly.

This Thesis would not have been in your hands without the financial support by the French Government, allocated through the French Embassy in Vietnam and Campus France.

Finally, my thanks go to my cousin living in Paris, his wife and his kids, who have become my second family in France. I deeply thank my beloved family whose unconditional love, support, unwavering confidence and patience with me have sustained me through this adventure.

# Contents

# General Introduction

## Motivation

*Horizontal inequality and poverty*

Inclusive growth cannot truly be attained if inequality is not tackled. Inequality, a broad concept, has been used vastly in different disciplines; searching this word in Google scholar gives us more than 3 million results. No one can deny the existence of inequality and the danger of it as a barrier to sustainable and inclusive development. Reducing inequality within and between countries has been put at heart of the United Nations' Sustainable Development Goals (SDG10): "*There is growing consensus that economic growth is not sufficient to reduce poverty if it is not inclusive and if it does not involve the three dimensions of sustainable development – economic, social and environmental… To reduce inequality, policies should be universal in principle, paying attention to the needs of disadvantaged and marginalized populations.*" Recently, a growing body of research has focused on horizontal inequalities (HIs), which refer to the gap between different groups in a society as categorized for example by race, ethnicity, or gender. Horizontal inequality, in many cases, is found to be persistent over a long period due to both, unequal access to different types of capital and different returns to capital accumulated. The persistence of HIs leads to individuals being stuck in poverty traps and obtaining lower outcomes in terms of human development indexes (related to for example, health, education, community participation, self-identify), exacerbating inequality of opportunities. Furthermore, a high level of HIs is positively correlated with a higher risk of conflict, and economic and political instability. Addressing HIs is also needed if the goal is to reduce individual or vertical inequality since vertical inequality includes both between and within group disparities.

The role of ethnic inequality as an important form of HIs has recently gained more attention in the literature, focusing especially on poverty rates among indigenous people as well as economic inequality among and between different ethnic groups. In a comprehensive review of poverty and socio-economic indicators among indigenous people, Hall and Patrinos (2012) highlighted common features of indigenous people across different countries: they are usually among the poorest and face many disadvantages. Another strand of literature has explored the negative association between ethnic fragmentation and economic development (Easterly and Levine, 1997; La Porta et al., 1999; Alesina et al., 2003). That inverse relationship is also documented in other studies considering different measures of ethnic diversity such as group polarization (Montalvo and Reynal-Querol, 2005) or ethnic segregation (Alesina and Zhuravskaya, 2011). Furthermore, recent work by Alesina et al. (2016) has shown the direct negative relationship between ethnic inequality and development. The authors emphasized the fact that is it economic inequality between different ethnic groups which is harmful to development rather than ethnic diversity per se. Their finding, using data from 173 countries, is in line with few other studies, such as Steward (2002) or Chua (2003), which provide further evidence from case-studies. Alesina et al. (2016) also highlighted the role of geographic endowments in shaping economic inequality between ethnic groups. Focusing on the mechanisms behind this inverse relationship between ethnic inequality and development, the literature has found links between ethnic inequality and inadequate public good provision (Baldwin and Huber, 2010; Alesina et al., 2014); between ethnic

inequality and lower government quality (Kyriacou, 2013); and between ethnic/racial inequality and the formation of social capital (Chantarat and Barret, 2011; Tesei, 2014).

*Why Vietnam?*

Vietnam has been one of the most cited success stories of economic growth and poverty reduction over the last three decades. This remarkable achievement can be attributed to the 'Doi moi', a broad set of national economic and political reforms starting in 1986 which have transformed Vietnam from one of the poorest countries into a lower middle-income one (Pimhidzai, 2018). Besides becoming a lower middle-income country in 2009, Vietnam has also achieved most of the Millennium Development Goals. Regarding poverty reduction, the poverty headcount plummeted from nearly 60 percent in the early 1990s to only 20.8 percent in 2010 and 9.8 percent in 2016 (World Bank, 2012; Pimhidzai, 2018). However, economic growth has not been fully inclusive. In terms of poverty, vulnerable households just above the poverty line are actually more likely to fall into poverty than before and it has become more difficult to reach the remaining poor due to a slow-down in poverty reduction and a rise in inequality. One of the biggest challenges to inclusive growth is ethnic poverty. In Vietnam, there are 53 ethnic minority groups, who represent less than 15 percent of the population but more than 50 percent of the poor in 2010 (World Bank, 2012). Furthermore, ethnic minority groups also perform worse on other welfare indicators, such as education, health, etc. The Country Social Analysis (CSA) report (World Bank, 2009) refers to this situation as a dilemma, given the fact that ethnic inequalities persist, despite a great amount of financial resources having been invested to lift up the situation of ethnic minority in the remote and mountainous areas and the fact that some ethnic minorities are living in parts of the country that are experiencing rapid economic growth. Concerns over inequality of opportunities between ethnic groups have also arisen, in particular, the ability to access education and health services (World Bank, 2014; Oxfam, 2017). Group-based inequality, if not addressed, will reduce not only economic growth but also achievements on other welfare indicators and might cause higher risk of instability. Understanding the causes of these persistent ethnic gaps and the potential of ethnic minorities is therefore a necessity to turn the country's fast economic growth into truly inclusive growth.

This introduction provides the reader with the general context of ethnic inequality in Vietnam and its evolution over time, emphasizes gaps in the literature on ethnic minority, and offers a conceptual framework that guides the research questions in this thesis. I also provide quotes extracted from my in-depth interviews with ethnic minorities living in Ha Giang, Ha Noi and Tay Ninh provinces, carried out in May, June and July 2015. The interviews provide insights from local ethnic minorities about their lives, their struggles and economic opportunities; they complement the quantitative data used in my thesis.

**Ethnic minorities in Vietnam: who are they and what do we know?**

Vietnam is a multi-ethnic country, with 54 ethnic groups, including 53 ethnic minority groups. The majority group are Kinh people, making up around 85 percent of the country's population. The remaining 53 ethnic groups are considered ethnic minorities in Vietnam. They mainly live in the mountainous and highland areas of Vietnam but are spread out from the North to the South (except for the Hoa/Chinese group who, together with the Kinh, mainly occupies the delta and coastal areas). Population size of the ethnic groups varies considerably. Among the ethnic minorities, the largest groups are the Tay, Thai, Muong, Hoa, Khmer, Nung with an average population of 1 million each, while other groups such as the Brau, Ro Mam and O Du only consist of several hundreds of members. In terms of language, there are five language families shared by the different groups. The Kinh language, or Vietnamese, which belongs to the Viet-Muong language group, is the national language and the second language of ethnic minority groups. The past century, with the Vietnam war, migration and government policies, has marked a significant change in ethnic minorities' lives. Minorities used to live in the upland meaning they were isolated from the Vietnamese who occupied the lowland and coastal land; today, however, the physical distance has been remarkably reduced. Also, due to the migration policies of the Government after 1975, Kinh were encouraged to resettle in the mountainous and less populated areas. As a result, many areas are now occupied by both Kinh and ethnic minorities.

The aforementioned ethnic poverty and inequality have been challenges for Vietnam due to the persistent and widening gaps in living conditions and poverty rates between the ethnic majority and minorities (World Bank, 2012). During the very first decade after the Doi moi reforms, from 1992 to 1998, the probability of escaping poverty was lower for ethnic minorities than majorities (Glewwe, Gragnolati and Zaman, 2002). Then, in the period of 1998 - 2010, per capita expenditures have grown at an average annual rate of 9.4 percent for the Kinh group and only 7.4 percent for ethnic minorities. The disparities are largest in some of the poorest and least accessible regions of Vietnam (World Bank, 2012). Even the fastest-growing minority households are growing more slowly than the average Kinh households. Between 2010 and 2014, 49 percent of ethnic majority that were in the bottom quintile of the income distribution moved to a higher income bracket, while the corresponding figure for other ethnic groups is only 19 percent (Oxfam, 2017). Using a biannual panel dataset in rural Vietnam from 2006 to 2014, Singhal and Beck (2015) also show that there is a significant difference between the Kinh households and the remaining ethnic groups and that this gap has not narrowed over time. Figure 1 provides a broad view of ethnic distribution in Vietnam and differences in spatial distribution of poverty rates between ethnic minorities and the majority.

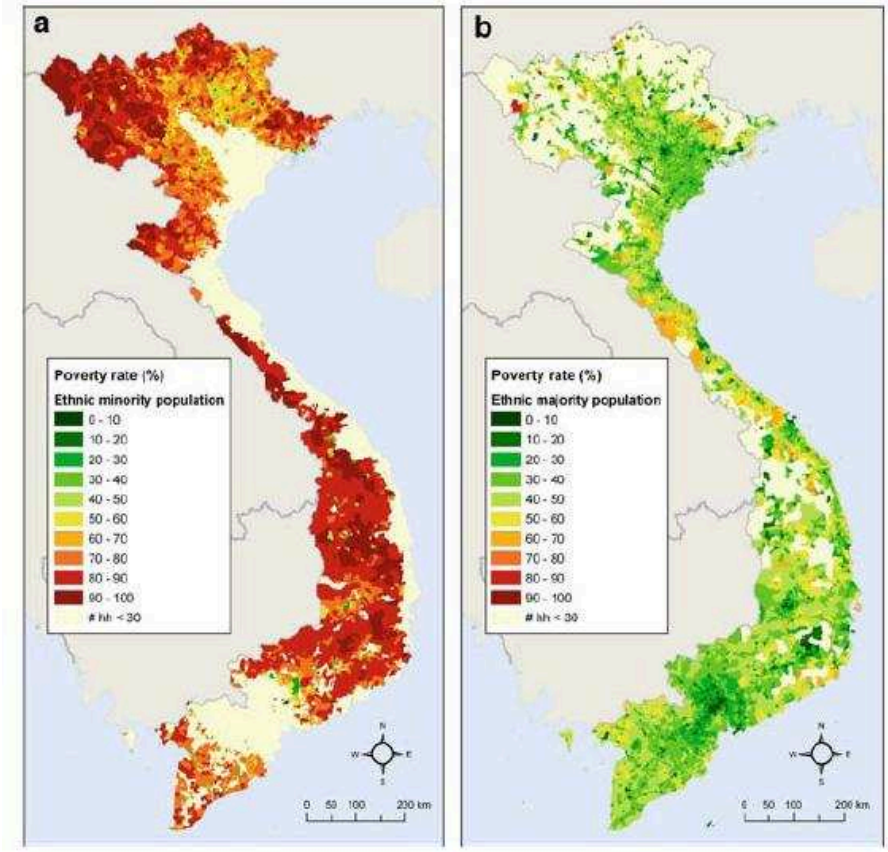



Figure 1. The spatial distribution of poverty rate for ethnic subgroups (extracted from Epprecht et al., 2011)

**Which factors could explain the ethnic gaps?**

Similar to the common root of ethnic inequality highlighted by Alesina et al. (2016), physical remoteness or geographical isolation is one the biggest disadvantages for ethnic minorities in Vietnam. The literature has shown the important role of geographical capital in determining welfare since it is associated with the quality of land, topography, access to basic infrastructure and public services, access to the market economy and to the political and economic centers (Epprecht et al., 2009). Spatial poverty traps (Ravallion and Jalan, 1997) indeed are a severe problem for the poor as the geographical disadvantages result in poor infrastructure and low returns to private endowments, hindering people living in such places from escaping poverty. This is the case in Vietnam and it has been one of the key components explaining the gap. However, the literature also shows that persistent ethnic inequality is not just attributable to physical accessibility (Baulch et al. 2007, 2012; van de Walle and Gunewardena, 2001).

Other key factors driving this persistent gap are the disparities between Kinh and non-Kinh in endowments and especially returns to endowments. Ethnic minorities indeed have fewer physical assets such as land holdings, access to credit, etc. and also fewer social assets such as education, health, and access to social services than Kinh. For instance, in 2006, only 6 percent of the Kinh household heads had no education compared to 23 percent of ethnic minority ones (World Bank, 2009). However, while the gap in physical distance has been addressed through infrastructure programs in rural and mountainous areas where ethnic minorities are concentrated, the more worrisome factor is that ethnic minorities have lower returns to their endowments. During the

period of 1993 – 2004, Baulch et al. (2012) find that half of the per capita expenditure gap by ethnicity is explained by ethnic differences in returns to endowments. In 2002, Pham and Reilly (2009) also show that returns to endowments explain about two third of the wage gap between ethnic minorities and the majority.

The underlying story of lower returns to endowments is rooted in several potential factors. It could be due to the social and cultural distance between the Kinh and non-Kinh or it could be explained by current and/or past discrimination (Baulch et al., 2007, 2012; van de Walle and Gunewardena, 2001; World Bank, 2009). For instance, provided with the same new technology for farming, an ethnic minority farmer might not use it as efficiently as a Kinh because of her different knowledge and customs. In the labor market, especially in the informal sector, non-Kinh could be discriminated against. A wage worker could be paid lower compared to the majority; a self-employed could face some discrimination from the demand side, for example, from her customers. Thus, the lower returns to endowments could be a consequence of both cultural differences and discrimination. Cultural and social distance are due to ethnic differences in language, social norms, traditional practices, etc. A recent study by Nguyen et al. (2017) finds that language is the main barrier of the ethnic minority's economic growth, leading to the increasing ethnic inequality. Unlike other studies, it shows that there is little evidence about the large explaining attribute of differences in returns to endowments. In addition, cultural and social distance also hampers social capital, especially in areas in which both Kinh and ethnic minorities reside.

> *"I got married to a Kinh man. In general, there is no discrimination. My brother in law also got married to a Tay like me so sometimes, my parents in law call us "bon dan toc" (you guys, ethnic minority) as a joke but it somehow implies a distinction between myself, my brother in law's wife and the rest of the family."*
>
> (Ms. H, Tay ethnic, Phong Quang commune, Vi Xuyen district, Ha Giang province)

> *"I rarely leave the village; I don't feel confident; I don't speak Vietnamese very well…"*
>
> (Ms. E, S'tieng ethnic, Minh Tam commune, Hon Quang district, Tay Ninh province)

Discrimination is indeed one disadvantage that ethnic minorities face, but it can come in subtle ways making it hard to measure and address. The CSA (World Bank, 2009) findings show that the majority still sees ethnic minorities as 'somewhat less developed' or even 'backward', thus, ethnic minorities need to be helped to 'catch up' with the Kinh people. As a result, government interventions based on the perception that ethnic minorities should be assimilated to the majority cannot be effective (further discussion is presented in the next paragraph). Furthermore, this long-lasting stereotype also has negative effects on how ethnic minorities feel and think about themselves, reducing their levels of confidence and self-esteem. As a consequence, it lowers their civic participation and reduces their capacity in having their own voice and power (World Bank, 2009). Indeed, historical discrimination could also trigger the differences in endowments between ethnic minority and majority groups today. The fact that ethnic minority households often have worse quality land or live in places with less access to markets is generally due to past (often far past) and cumulative discrimination (van de Walle and Gunewardena, 2001).

> *"Ethnic minorities don't know how to make a living."* – a cadre of the Committee for Ethnic Minority Affairs
>
> *"Minorities don't know how to use credit efficiently"* – a Social Policy Bank cadre
>
> *"Minorities don't have the will to get ahead"* – a District People's Committee member
>
> (extracted from the Country Social Analysis: Ethnicity and Development in Vietnam, World Bank 2009)

Lastly, numerous programs sponsored by the government of Vietnam and funded by both national and international organizations have been implemented to alleviate poverty and improve the welfare of ethnic minorities. However, the effectiveness of those programs has been questioned (Dang, 2012) due to limited data on costs and benefits, or proper impact evaluations. Some improvements in poverty reduction and income growth have been observed for Tay, Nung, Dao, and H'mong groups but less so among others (Nguyen et al., 2013). These unequal effects are attributed to the fact that some projects and policies do not match the culture, specific needs and conditions of some ethnic minority groups (van de Walle and Gunewardena, 2001; Nguyen et al., 2013). This issue comes from such interventions being designed based on the Kinh's conditions, behaviors and cultures, which are not necessarily applicable to non-Kinh. Besides, as the policies target ethnic minority areas rather than ethnic minority households themselves, a problem of elites/Kinh capturing benefits also exists. Given the isolation in living areas of ethnic minority, policies designed by the higher level and by people who live far away and do not share the same geographical, cultural and economic backgrounds with the policy's beneficiaries, is an issue. In addition, the internal diversity within ethnic minorities in terms of geography and culture is a challenge in tailoring the suitable programs. "*Indigenous knowledge and local practice should be informing the work of scientists and researchers should clearly be taking local knowledge into account when designing new technologies to make them locally suitable. Researchers and extension agents also need to gather information on whether what they develop and promote actually works or does not work – a feedback loop that is currently missing in the management of research and extension. This information would lead to improvements in the 'models' and make them more suitable for local conditions*" (Swinkels and Turk, 2006).

**What lacks in the literature? Data and research limitations**

The first and foremost difficulty in studying ethnic inequality is data limitation. In spite of a growing number of studies on ethnic minority poverty in Vietnam, most of them can only capture the situation of some main ethnic groups due to a lack of data and a limited number of observations. Most of the previous quantitative works on ethnic issues in Vietnam have focused on the separation between Kinh and Non-Kinh or between Kinh/Hoa and the rest (van de Walle and Gunewardena, 2001; Baulch et al., 2007) since the Hoa also reside mainly in the delta areas and they are as well-off as the Kinh. More recent papers have considered 8 ethnic groups: Kinh & Hoa, Kho-me & Cham, Thai, Tay, Muong and Nung, Other Northern minorities, Central Highland minorities, Others (Baulch et al., 2012). Thus, it limits the explaining power of studies when different ethnic groups are aggregated as one ethnic minority group or grouped as 'others'. These 53 ethnic minority groups are diverse in terms of assimilation and economic achievements. Thus, it is important to disaggregate the ethnic minority groups as much as possible to better understand their situation and design relevant policies accordingly. Regarding ethnic categorization, the definition of minority and majority need to be re-considered. Across the country, Kinh is the ethnic majority but in a given area, one ethnic minority or other ethnic minorities are, in fact, the majority. Accordingly, each agent would be expected to act differently depending on their relative position. Besides, local context in terms of ethnic diversity is also a

key factor affecting development as suggested in the literature but addressed in only few studies on ethnic inequality in Vietnam (Nguyen et al., 2017). Another concern about data limitation in explaining the ethnic gaps by endowments and returns to endowments is measurement. Quality of endowments (land, infrastructure, etc.) is hard to capture, thus, it might lead to some biases in estimating the returns to ethnic minority and majority's characteristics (Nguyen et al., 2017). Lastly, by reviewing briefly the literature on ethnic inequality in Vietnam, I find that ethnic gaps have been investigated in relation mainly to four types of capital, namely natural, human, financial and physical capital but little is known about social capital and social networks. Since language has been found as a barrier to ethnic development, Nguyen et al. (2017) also suggest the necessity for further studies on the effects of social network and community effects in explaining ethnic inequality in Vietnam.

# Chapter summary and contextualization

This thesis combines three essays on poverty and ethnic inequality in Vietnam. These three chapters can be approached independently of each other but also complementarily as they weave a coherent and comprehensive picture of the drawbacks in the process of poverty reduction and the roots of ethnic gaps underlying horizontal inequalities in Vietnam. Being aware of the literature gap, in this thesis, I aim to tackle the gap and overcome some limitations in previous studies through three different angles. The first chapter investigates if the community participation approach is implemented in poverty reduction programs targeted at mountainous areas and ethnic minorities, and evaluates the effect of community participation on households' economic and welfare outcomes. The second chapter examines whether gender and ethnic earnings gaps exist in the Vietnamese labor market and how much of these gaps could be explained by differences in endowments and returns to endowments. Lastly, the third chapter explores the disparities in risk sharing networks between the ethnic minority and majority and highlights the importance of cultural and social distance in social network formation. The remainder of this introduction provides the summary of each chapter and its context motivating the research questions.

I approach the issue of ethnic inequality in Vietnam through three different mechanisms as follows:

Chapter 1: Poverty reduction programs

Chapter 2: Labor markets

Chapter 3: Social networks

**Chapter I** deals with the question of community participation in a poverty reduction program in rural areas where mostly ethnic minorities reside. It gives a broad assessment of project implementation with a focus on the participatory approach.

Since geographic disadvantage is the first and foremost factor considered as explaining ethnic inequality, the Government of Vietnam has invested in commune infrastructure as an instrument to narrow the gap. Many projects have been carried out, but their impacts are still questioned given the fact that ethnic gaps have not been reduced. Thus, the lack or insufficient benefits of the projects raises concerns about how these infrastructures are provided and to what extent local communities, ethnic minorities in particular, are involved in these projects as they are the targeted beneficiaries of these projects.

Community-Driven Development (CDD) has become a key strategy used by governments and development assistance programs. CDD is expected to develop projects and programs that are sustainable and responsive to local priorities, to empower local communities to manage and govern their own development programs, and more effectively target poor and vulnerable groups. However, the empirical evidence of its effectiveness in achieving these objectives is mixed. Theoretically, participatory methods can improve outcomes, but under some circumstances, participation is not effective. While many CDD initiatives may be promising, their impact is often limited (Gillespie, 2004). Vietnam provides an interesting case to test the effectiveness of this approach. On the one hand, Vietnam has been one of the highest recipients of IDA loans from the World Bank *"In 2008, new commitments broke through the one billion dollar*

*mark… Vietnam was now the number one recipient of IDA loans, ahead of India and Bangladesh"* (Cling et al., 2009). All development initiatives such as CDD and community participation have been rapidly adopted as part of the poverty reduction programs in Vietnam, not only to practice poverty alleviation effectively but also to attract more development loans. On the other hand, Vietnam is a ruled by one Communist Party with a long history of a centralized system and a command economy while the civil society does not legally exist (Cling et al., 2009). This might not be a setting in which the CDD approach could flourish.

Despite the strong emphasis on theoretical arguments on the benefits of CDD approach and the increasing application of this approach in Vietnam's development projects, little evidence is known about how this CDD approach is actually implemented in the field. Culas et al.'s study (2015) is one of the few exceptions. Using both quantitative and anthropological data, they show that participation indicators are improved, but remain problematic in practice. In this study, their focus is on the governance component of the "Poverty Reduction Program 135" using the baseline dataset, only. The authors are thus not able to assess the benefits of household participation on welfare outcomes. To fill this gap, my paper aims at providing an assessment of the implementation of CDD in the context of Vietnam and examining the effect of household participation. The ambition of this chapter is threefold: (i) to examine how CDD has been implemented in Vietnam's poverty reduction program; (ii) to measure the effect of household participation on economic and welfare outcomes and (iii) to evaluate if there is any significant association between better CDD implementation and communes' outcomes.

This chapter takes advantage of a unique panel data between 2007 and 2012 from a survey carried out in the most disadvantaged and mountainous areas of Vietnam under the "Poverty Reduction Program 135". It provides an analysis of this participatory approach at both the household and commune levels. I assess the effect of household participation on both final outcomes including income, productive assets, durables assets as well as a series of other welfare indicators or intermediate outcomes. I also examine whether there is any heterogeneity effect of participation for the poor and ethnic minorities. Results show that active participation has a large effect on income growth. By examining the effect of participation on intermediate outcomes, I find that participating households can get access to better resources, in particular, information and communication, which could be the channels leading to their better final outcomes. Regarding the heterogeneity of participation effect, I find higher benefit for the poor and ethnic minorities on intermediate outcomes. At the commune level, there is a positive association between CDD implementation and economic outcomes. Evidences from this chapter show that active participation could help the poor and ethnic minority to benefit more from development projects.

**Chapter II** provides a comprehensive view on ethnic earnings gaps in Vietnam using the Labor Force Survey along with gender earnings gaps as a comparison and supplement.

A job brings a person an earning, thus, the issue of ethnic poverty should be closely investigated from a labor market and employment perspective. Vietnam's successful story of economic growth cannot be achieved without a deep transition from a 'centrally planned economy' towards a 'socialist-oriented market economy'. The labor market has been profoundly changed, with employment shifting from agriculture to non-agricultural sector, the downsizing of jobs in State-owned enterprises and remarkably fast development of household business and the informal sector. First, this transformation is a challenge for women and ethnic minorities who are more dependent on the agricultural sector than men and ethnic majority counterparts. Similar disadvantage could arise due to the downsizing of State-owned enterprises. Second, when the

economy shifts from the agricultural sector to the non-agricultural sector, the share of wage workers increases and it is seen as a structural phenomenon in Vietnam. It has increased by 4 percent each year between 2002 and 2006 and reduced to 2 percent increase in the period of 2007 – 2012 (Oudin et al., 2014). Wage workers are considered to be better protected by legal regulations, thanks to the establishment of Labor Code, etc. Women receive more privileges at work such as fully-paid maternity leave or exemption from unilateral termination of their contract during pregnancy. Outside the public sector, however, with the autonomy in recruitment, employers could favor men over women to avoid these privileges (Pham and Reilly, 2007). Another remarkable change is the emergence of the informal sector. This sector has attracted around 25 percent of the total employment or half of the jobs outside agriculture; meanwhile, this informal sector is characterized by low earnings and precarious working conditions, which are especially detrimental to women (Cling et al., 2014). When the farm and public sector jobs are reduced, the informal sector might absorb the surplus of labor supply, in particular women and ethnic minority. Thus, these sectoral bring challenges for women and ethnic minorities.

For ethnic minority groups, the changes in their labor market participation rates are associated with a number of government policies aiming at creating more job opportunities for ethnic minorities and motivating them to participate in the non-farm sector. With the man-power 'quotas' allocation policy, part of the central planning regime, jobs were guaranteed in the public sector for ethnic minority graduates from secondary and tertiary school (Pham and Reilly, 2009). Graduates from the majority are encouraged to work in the mountainous areas while minorities are provided jobs in the public sector in urban areas. Education for ethnic minorities is also promoted with different kinds of support/assistance for ethnic minority students going to public colleges and universities. Therefore, thanks to the stronger investment in human capital and different policies promoting ethnic minorities' participation in non-farm jobs, especially in the public sector, it is expected that ethnic minorities have more opportunities in labor market.

> *"During my bachelor at the University of Social Sciences and Humanities, I did not have to pay the tuition fees because I'm Thai ethnic. Besides, I received a support of 700 thousand VND per month for living expenses from the Government for being an ethnic minority student."*
>
> (Ms. S., Thai ethnic, Institute of Anthropology)

However, the CSA (World Bank, 2009) also reports constraints for ethnic minorities in the labor market due to cultural factors in market interactions, schooling, etc. Their cultural norms make them less active in making money out of an economic transaction, for instance, they are less likely to charge interest rate from kin or neighbors. Women in ethnic minorities groups face even more barriers in work because cultural norms disfavor women and their position in decision making is secondary. Thus, in this context, the question is whether ethnic women face more disfavors in the labor market in comparison with ethnic men, Kinh women, and Kinh men.

> *"The buyer goes to our village to buy rice, so we do not need to move. I'm not sure if the price is fair or not."*
>
> (Mr. K. , S'tieng ethnic, Minh Tam commune, Hon Quang district, Tay Ninh province)

> *"If there is a meeting in the village, my husband goes to the meeting."*
>
> (Ms. N. , S'tieng ethnic, Minh Tam commune, Hon Quang district, Tay Ninh province)

This chapter uses the 2007 Labor Force Survey to estimate and analyze the gender and ethnic earnings gaps in Vietnam. Previous researches have shed light on the potential for discrimination against women and ethnic minorities. Their studies focused mostly on wage employment and disregarded the informal sector. However, in Vietnam's labor market, women's participation in the informal sector is substantial and a majority of ethnic minority workers are engaged in non-wage occupations. Previous studies were thus severely limited in analyzing the gender and ethnic gaps separately; furthermore, they failed to disaggregate the different ethnic groups and consider the subgroup of ethnic women, in particular. I overcome these aforementioned issues by including the self-employed and agricultural workers in estimating the gender and ethnic earnings gaps and investigate further the heterogeneity in the gaps among different ethnic groups.

This chapter shows that the earnings gap between ethnic groups is larger than the earnings gap by gender and results are heterogeneous amongst different ethnic minority groups. After employing a multinomial logit model to correct for selection, I find that the gaps depend significantly on the sectoral choice. Selection is thus playing a key role in the Vietnamese labor market. Using decomposition techniques, I show that differences in sectoral allocation play an important role in explaining earnings gaps by ethnicity while gender gaps are observed mainly within sectors. Lastly, unexplained earnings gaps still remain for both ethnic and gender gaps, even after correcting for selection.

**Chapter III** focuses on the differences in risk sharing networks between ethnic minorities and majority groups and relies on the concept of ethnic homophily in link formation to explain the mechanisms leading to those disparities.

This chapter is motivated by the fact that risk and the necessity to cope with it is a pervasive feature of rural economic life, particularly in Vietnam's mountainous areas, where ethnic minorities mostly live. There, the prevalence of risk, and its role in the persistence of poverty, is accentuated by climate change, by difficulties to cultivate fragile uplands, and by high transport costs, combined with low access to basic infrastructure, public services and formal insurance. In this context, risk sharing arrangement, in which households share the risk with others through transfers in gift, money and/or labor, is a dominant mean to cope with shocks for the rural poor (Rockenbauch and Sakdapolrak, 2017).

An important question is then with whom people choose to connect and share the risk. Ethnicity is potentially an important factor affecting how a risk sharing network is formed (Grimard, 1997). If so, ethnic minority groups may differ from the Kinh majority in the characteristics of their risk sharing network, and thus in their capacity to mitigate the effect of shocks. This question of risk sharing network formation from an ethnicity perspective is particularly relevant in the context of Vietnam but it has not been studied yet. As mentioned above, social and cultural distance between ethnicities exists.

This chapter mobilizes various datasets for identification and measures of social networks and outcomes. I use the 2008-2016 Vietnamese Access to rural Resources Household Survey (VARHS), matched with the Vietnam Census of 2009 to extract information on ethnic composition and occupation at the commune level, and with Geo-data to characterize households' environmental contexts. The methodology relies on the estimation of a series of models explaining risk sharing network characteristics (size, ethnic similarity, occupational and geographical similarities), introducing step by step various sets of covariates that account for socio-economic characteristics, geographical environments, demographic and local distribution of ethnic groups, and the ethnicity of the household head.

I show that ethnic minority groups are embedded in smaller risk sharing networks, which implies their lower capacity to pool risks. In addition, their networks are less diversified or, in other words, the people within ethnic minority network are more similar in terms of occupation and location than within Kinh networks. Being connected with people with the same economic activity and living in the same village is expected to be less efficient to cope with covariant shocks. Indeed, covariant shocks, by definition, affect people in the same geographical area or sector, by contrast with idiosyncratic shocks which are peculiar to an individual. Covariant shocks are most prevalent in the rural areas of Vietnam.

First, I find evidence that the gaps between Kinh and non-Kinh risk sharing networks are at least partly driven by the social distance that exist between the Kinh and other ethnic groups. Indeed, Kinh households are found to have higher preferences for large networks and for inbreeding homophily than other groups and therefore tend to exclude ethnic minorities from their networks. I show that this allows Kinh to form networks that are more diverse in terms of occupation. Beyond this result, I find that social network dissimilarities also hinge on the demographic imbalance between ethnic groups, as well as on differences in wealth, education, and in geographic and institutional commune-level characteristics. Ethnic inequality is thus partly rooted in the cultural and social distances that exist between ethnic groups in Vietnam.

This chapter provides evidence from a developing country on the role played by ethnicity in the formation of risk sharing networks and distinguishes between various mechanisms at play. Better understanding the differences in risk sharing networks between ethnic groups is likely to provide valuable information to address the ethnic gap which is one of the most challenging issues in the context of Vietnam.

**References**


Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. 2003. 'Fractionalization'. *Journal of Economic Growth*, 8(2): 155-194.

Alesina, A., and E. Zhuravskaya. 2011. 'Segregation and the Quality of Government in a Cross Section of Countries.' *American Economic Review*, 101(5): 1872-911.

Alesina, A., S. Michalopoulos, and E. Papaioannou. 2016. 'Ethnic Inequality.' *Journal of Political Economy,* 124 (2): 428-88.

Baldwin, K. And J. Huber. 2010. 'Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision.' *American Political Science Review*, 104(4): 644-62.

Baulch, B., T. T. K. Chuyen, D. Haughton, and J. Haughton. 2007. 'Ethnic minority development in Vietnam.' *Journal of Development Studies*, 43(7): 1151-1176.

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.

Chantarat, S., and C. B. Barrett. 2012. 'Social network capital, economic mobility and poverty traps'. *Journal of Economic Inequality*, 10: 299-342.

Chua, A. 2003. *World on Fire: How Exporting Free-Market Democracy Breds Ethnic Hatred and Global Instability*. New York: Doubleday.

Cling, J.-P., M. Razafindrakoro, and F. Roubaud. 2009. 'Desperately seeking model countries: The World Bank in Vietnam.' Document de Travail DIAL DT/2009-04. Paris: DIAL.

Cling, J-P., Razafindrakoro, M., and Roubaud, F. 2014. 'Informality, crisis and public policies in Vietnam'. In: Cling, J.P. (ed.), Lagrée, S. (ed.), Razafindrakoto, M. (ed.), & Roubaud, F. (ed.). (2014) *The informal economy in developing countries*. Oxon : Routledge.

Culas, C. , M. Razafindrakoto, and F. Roubaud. 2015. 'Ownership and Participatory processes: from global motto to local challenges. The case of a key poverty reduction program in ethnic minority areas of Vietnam.' In *Local Politics, Global Impacts: Steps to a Multi-Disciplinary Analysis of Scales*. O. Charnoz, V. Diaz Pedregal, A. L. Kolata (eds.). Surrey, UK: Ashgate Press: 223-250.

Dang, A. 2012. 'A Widening Poverty Gap for Ethnic Minorities'. In *Indigenous Peoples, Poverty and Development*. G. Hall and H. Patrinos (eds.). Chapter 8. Cambridge, NY: Cambridge University Press.

Easterly, W., and R. Levine. 1997. 'Africa's Growth Tragedy: Policies and Ethnic Divisions.' *The Quarterly Journal of Economics,* 112(4): 1203-1250.

Epprecht, M., D. Müller, and N. Minot. 2009. 'How remote are Vietnam's ethnic minorities? An analysis of spatial patterns of poverty and inequality.' *The Annals of Regional Science*, 46(2): 349-368.

Gillespie, S. 2003. 'Scaling up community driven development: a synthesis of experience.' Washington, D.C.: World Bank.

Glewwe, P., M. Gragnolati, and H. Zaman,. 2002. 'Who Gained from Vietnam' Boom in the 1990s?' *Economic Development and Cultural Change* 50(4): 773-92.

Grimard, F. 1997. 'Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire.' *Journal of Development Economics*, 53(2): 391-422.

Hall, G. H., and H. A. Patrinos (eds.). 2012. *Indigenous Peoples, Poverty, and Development*. Cambridge, NY: Cambridge University Press.

Kyriacou, A. P. 2013. 'Ethnic Group Inequalities and Governance: Evidence from Developing Countries.' *Kyklos,* 66(1): 78-101.

Montalvo, G., and M. Reynal-Querol. 2005. 'Ethnic Polarization, Potential Conflict, and Civil Wars.' *American Economic Review* 95(3): 796-816.

Nguyen, C., T. Phung, T. Phung, T., N. Vu, and D. Westbrook. 2012. 'The Impact of A National Poverty Reduction Program on Ethnic Minorities in Vietnam: The Lens of Baseline and End Line Surveys.' MPRA Paper, 50477.

Nguyen, H.-T.-M., T. Kompas, T. Breusch, and M. B. Ward. 2017. 'Language, Mixed Communes, and Infrastructure: Sources of Inequality and Ethnic Minorities in Vietnam.' *World Development*, 96: 145-162.

Oudin, X., L. Pasquier-Doumer, M. T. Pham, F. Roubaud, and H. D. Vu. 2014. 'Adjustment of the Vietnamese Labour Market in Time of Economic fluctuations and Structural Changes.' DIAL Document de Travail, DT/2014-04. Paris: DIAL.

Oxfam. 2017. Even it Up: how to tackle inequality in Vietnam. Oxfam briefing paper.

Pham, H. T., and B. Reilly. 2009. 'Ethnic wage inequality in Vietnam.' *International Journal of Manpower*, 30(3): 192-219.

Pham, T., and B. Reilly. 2007. ' The gender pay gap in Vietnam, 1993–2002: A quantile regression approach.' *Journal of Asian Economics*, 18(5): 775-808.

Ravallion, M. and Jalan, J. 1999. 'Spatial poverty traps?'. Policy Research Working Paper Series, 1862. Washington, D.C.: World Bank.

Rockenbauch, T. and P. Sakdapolrak. 2017. 'Social networks and the resilience of rural communities in the Global South: a critical review and conceptual reflection.' *Ecology and Society*, 22(1): 10.

Singhal, S., and U. Beck. 2015. 'Ethnic Disadvantage In Vietnam: Evidence Using Panel Data.' UNU-WIDER Working Paper 2015/97.

Stewart, F. 2002. 'Horizontal Inequalities: A Neglected Dimension of Development.' UNU World Institute for Development Economics Research, WIDER Annual Lecture 5.

Swinkels, R. and Turk, C. 2006. 'Explaining Ethnic Minority Poverty in Vietnam: a summary of recent trends and current challenges.' Draft Background paper for CEM/MPI meeting on Ethnic Minority Poverty. Hanoi: World Bank.

Tesei, A. 2014. 'Trust, Racial Fragmentation, and Income Inequality: New Evidence from the US.' CESifo Working Paper Series, 4718. van de Walle, D., and D. Gunewardena. 2001. Souces of Ethnic Inequality in Vietnam. *Journal of Development Economics*, 65 (1): 177–207.

World Bank. 2009. 'Country Social Analysis : Ethnicity and Development in Vietnam - Main Report.' Washington, D.C.:World Bank.

World Bank. 2012. 'Well Begun, Not Yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges.' Hanoi: World Bank.

World Bank. 2014. 'Taking Stock, July 2014: An Update on Vietnam's Recent Economic Developments.' Hanoi: World Bank.

World Bank. 2018. 'Climbing the ladder : poverty reduction and shared prosperity in Vietnam.' Washington, D.C. : World Bank.

# Chapter 1

# Assessing Community-Driven Development approach in poverty reduction programs in Vietnam

**Abstract**

Community-Driven Development (CDD) has been put forward as one of the key principles in the development of public policies over the last 20 years, and it has been adopted in many poverty reduction programs at both global and local levels. We assess the implementation of this approach in poverty reduction programs in rural Vietnam by using a unique panel data over the period of 2007 – 2012. Results show that active participation has a large effect on income growth. By examining the effect of participation on intermediate outcomes, we find that participating households can get access to better resources, in particular, information and communication, which could be the channels leading to their final better outcomes. Regarding the heterogeneity of participation effect, we find higher benefit for the poor and ethnic minorities on intermediate outcomes. At the commune level, there is a positive association between CDD implementation and economic outcomes. Evidences from this chapter show that active participation could help the poor and ethnic minority to benefit more from development projects.

*Key words: Vietnam, ethnic minority, poverty*
*JEL classification: I30, I32, J15*

Acknowledgements

The current version of this paper has benefited from valuable comments at the 8th Vietnam Economist Annual Meeting in Thai Nguyen, Vietnam and research assistance from Rawane Yaser, Universite Paris 1.

# 1 Introduction

Poverty reduction is one of the most important targets in all developing countries. There have been many achievements in poverty reduction all over the world. However, there is still scope to improve the efficiency of such programs as it is seen that those programs have often bypassed or ignored community participation (Chebil and Haque, 2003). To address this limitation, international donors like the World Bank (WB), the International Monetary Fund (IMF) and governments of developing countries have been giving priority to poverty reduction strategies with a focus on governance, ownership and participatory approaches. Thus, these key principles have been proposed in the new international poverty reduction strategies (PRSP initiatives, MDG, HIPC, IADM, etc.) with the expectation of contributing to the effectiveness of poverty reduction practices (Cling et al., 2003). While there is wide consensus about the importance of those principles, the key question is about their effectiveness when being applied in specific situations in the field. Is it optimal to put local community and the poor in the driver seat? What are the necessary conditions to promote the role of beneficiary population in poverty reduction through using these approaches?

The community-driven development (CDD) has become a key strategy used by governments and development assistance programs due to its potential to develop projects and programs that are sustainable and responsive to local priorities, to empower local communities to manage and govern their own development programs, and more effectively target poor and vulnerable groups. However, the empirical evidence of its effectiveness in achieving these objectives is mixed. Theoretically, participatory methods can improve outcomes, but under some given circumstances, participation is not effective. While many CDD initiatives may be successful, their impact is often limited (Gillespie, 2004) and indeed, its benefits are associated strongly with local context.

Vietnam, a successful model of economic reform and poverty alleviation, has adopted the CDD approach in its development projects, following the global requirements of community participation. Besides, in 1998, a Decree on Grassroots Democracy has established a new governance framework for Vietnam. It aims at including all citizens in a more deliberative and democratic process of local governance, extending downward accountability, voice and empowerment of the citizens and limiting corruption. Despite the strong emphasis in theoretical arguments on the benefits of CDD approach, the increasing application of this approach in Vietnam as well as the country's political changes responsive to CDD approach, little evidence is known about how this approach is actually implemented in the field. To the best of our knowledge, no study deals with the question of its benefit on welfare outcomes. To fill the gap, this paper aims at providing an assessment on the implementation of the CDD in the context of Vietnam and examining the effect of household participation. The ambition of this chapter is threefold: (i) to examine how CDD has been implemented in Vietnam's poverty reduction program; (ii) to measure the effect of household participation on economic and welfare outcomes and (iii) to see if there is any significant association between better CDD implementation and communes' outcomes.

Taking advantage of a panel data covering around 4,000 households in almost 400 communes under the Poverty Reduction Program, we provide an analysis of participatory approach at both household and commune levels. We assess the effect of household participation on income, productive assets, durables assets as well as a series of welfare/intermediate outcomes. We also examine whether there is any heterogeneity effect of participation for the poor and ethnic minorities. Results show that only active participation has a large effect on income growth. By examining the effect of participation on intermediate outcomes, we find that participating

households can get access to better resources, information and communication in particular, which could be the channels leading to better final outcomes. Regarding the heterogeneity of the economic effect, we find no higher effect for the poor or ethnic minorities but some intermediate benefits. Regarding the CDD approach at commune level, there is a positive association between CDD implementation and commune's economic outcomes. Nevertheless, our study encounters some limitations. The first caveat of this paper is that we are only able to examine the effect of participation on household's outcomes in general and not to attribute the impact to a specific project, thus, the channels of the participation effect are not easy to directly interpret. We try to overcome this limitation by looking at different intermediate outcomes in order to provide plausible channels of the final effects. The second caveat is related to the Vietnam's history of a command economy where communes' authorities were given quotas and objectives that need to be met, thus, the participation indicator would be questioned as it implies an integral part of how the State and local authority mobilize and control citizens. This implicit nature of participation in Vietnam may bias the effect of participation, thus, we interpret the results with cautions and some discussions.

This paper is organized as follows: section 2 presents literature review, the context of Vietnam and motivation for this paper. Section 3 describes our data and methodology. Section 4 presents descriptive statistics. Section 5 details the effect of participation on final and intermediate outcomes as well as the effect of CDD approach implementation at commune level. Section 6 concludes.

## 2 Literature review

### 2.1 CDD and participation approach

By recognizing the limits of traditional approaches as well as the importance of participation in development, international aid organizations, multilateral organization and national government has shifted their interest into the community participation. It can noticeably be seen in the changes of WB/IMFs' PRSP, moving from a "*top-down*" to more of a "*bottom-up*" approach in tackling the formulation of national policies and strategies for poverty reduction. Since the 1990s, the WB's Community Driven Development (CDD) initiatives have been spread out and the CDD operates on "the principles of local empowerment, participatory governance, demand responsiveness, administrative autonomy, greater downward accountability, and enhanced local capacity" (Wassenich and Whiteside, 2004).

CDD is an approach to poverty alleviation, and the provision of infrastructure and services, that is rooted in the idea that development often works better when control over decisions and resources are handed over to local communities. According to World Bank definitions, community driven development depends on building partnerships of different kinds between community-based organizations, non-governmental organizations, and public and private sector service agencies (Shanks et al., 2003). CDD gives control of decisions and resources to community groups and treats the targeted population as partners in the development process, building on their institutions and resources (Dongier et al., 2003). The CDD approach is expected to involve stakeholders in local development investment and to empower communities to manage their own development initiatives. It is thus widely assumed that participation improves development outcomes, benefiting citizens more than operations that do not engage

citizens. CDD has the potential to make poverty reduction efforts more responsive to demands, more inclusive and more effective than centrally led programs (Dongier et al., 2003).

First, since CDD can make services more responsive to demands, it can enhance sustainability. Community-developed facilities such as health centers, schools, and water supply systems tend to have higher utilizations rates and are better maintained than when investment decisions are made by actors outside the community. Second, community management of development investments usually results in lower costs and more productively employed assets. These benefits can be shown in different sectors and services such as education, microfinance, and infrastructure. Finally, well-designed CDD programs are inclusive of poor and vulnerable groups, build positive social capital, and give them greater voice both in their community and with local government entities. It can provide voice and empowerment to groups that are typically excluded from the development process. It thus promotes transparency and accountability as well as control over decisions which give communities the opportunity to build social capital (defined as the ability of individuals to secure benefits as a result of membership in social networks) by expanding the depth and range of their networks.

However, despite all its potential benefits mentioned above, CDD also presents risks if not appropriately implemented. For this reason, a first branch of literature concerns the implementation of the CDD approach in different development projects and assesses its impact at the program level. CDD approach has become a strategy used by both government and development assistance programs. The impact of these programs has been assessed through different studies. General papers examine the overall impact of participation using evidence from different countries. Mansuri and Rao (2013) conclude that the poor tend to benefit less from participatory processes. "*Local participation does not work when it is merely the ad hoc, myopically directed, creation of a project. It works when it has teeth, when it builds on organic movements, when it is facilitated by a responsive center, when it is adequately and sustainably funded, and when interventions are conditioned by a culture of learning by doing.*"

White et al. (2018) synthesize evidence on 23 community-driven development programs to examine the impact of these programs. They conclude that CDD programs have made a substantial contribution to improving the quantity of small-scale infrastructure, but a weak effect on health outcomes and a very significant effect on education and other welfare outcomes. Evidence from 8 programs shows WASH community projects have a significant impact on improving access to improved water and sanitation. Regarding the impact on social cohesion and governance, a meta-analysis of CDD program effects shows that they have had no impact on social cohesion, decentralization or governance, which is due to the CDD program design that may favor the elite or to the lack of program rules and facilitation to encourage participation.

Although community participation has been widely promoted in development projects, empirical evidence of the effectiveness of CDD, in targeting poor and vulnerable groups and empowering local communities, is mixed (Mansuri and Rao, 2013). Khwaja (2004) observes that projects managed by communities were more sustainable than those managed by local governments because of better maintenance of the assets and infrastructure created by the project. Greater community participation in nontechnical decisions is associated with higher project outcomes whereas the opposite holds for community participation in technical decisions.

Arcand and Bassole (2007) observe that access to clean water and health services increased for poor families in a CDD project area in Senegal. First, they study the impact of treatment by the program on the accessibility of basic services, household expenditures and child anthropometrics, using a quasi-experimental approach. Second, using instrumental variable, they estimate the impact of completed projects on the household and child response variables, which allows them

to assess the magnitude of the impact of 'treatment on the treated'. Third, they use IV methods to estimate the impact of completed projects within geographical units that eventually get treated by the program. The results suggest that CDD infrastructure programs can improve the nutritional status of children in poor households. CDD would thus improve targeting because projects make better use of local knowledge to define and identify the targeted groups. However, a different view argues that in heterogeneous communities with high social inequality, the performance of CDD projects in targeting can be worse than that of externally managed programs.

Nkonya et al. (2012) use difference-in-difference, and propensity score matching approaches to evaluate the impacts of a CDD project, Fadama II, on household income and acquisition of productive assets in Nigeria. They find that the project succeeded in targeting the poor and women farmers in its productive asset acquisition component. The results show that participation in the project increased the income of beneficiaries by about 40-60 percent.

Narayan (1995) examines how final outcomes were affected by the quality of outcomes at each of the stages of a project cycle (design, implementation, construction and maintenance). They test the contribution of participation to the effectiveness and capacity building aspects of rural water supply projects To explore the path through which beneficiary participation affects outcomes, the proximate determinants were divided into two broad categories: institutional outputs (design and implementation of a project), and physical outputs (construction and maintenance of water systems). They conclude that for maximum benefits, beneficiary participation needs to be viewed as a long-term process that necessitates involvement of users from the beginning of a project to its end, at all stages of a project cycle, revealing thus the importance of implementing a CDD approach and not just the carry-out of certain mechanical tasks by the beneficiaries.

On the other hand, some evidences from randomized experiments have provided a different view. Banerjee et al. (2010) find no effect of giving local committees the tools to evaluate student performance and monitor teachers. They argue that providing information on the status of education and the institutions of participation alone is not sufficient to encourage beneficiary involvement in public schools and that to improve final outcomes, fostering and channelling local action may be the most effective.

Drawing up upon this literature and findings, we can conclude that CDD should be viewed as a long-term process beyond the project. The empirical evidence on the effectiveness of CDD approach remains mixed, with some papers arguing for the positive impact of CDD on outcomes and other papers find no effect on outcomes, highlighting the idea that the results seem to depend on the details of the intervention and the local contexts.

*Beneficiary (household) participation*

As a specific component of CDD, household participation is promoted due the benefits that household can derive from directly and indirectly. To understand the potential effect of household participation, we locate our review of motives for household participation in a broader scope than within community participation under CDD or within a specific project. First, households participate in community work to enhance local public goods from which they can derive direct benefit. Second, household participate in community work when they have incentive to create or strengthen social capital, particularly for the new comers in a village. Related to the social capital and social network perspective, household may participate in community work in order to share information about income generation activities, including the adoption of new technologies (Bandiera and Rasul, 2006; Conley and Udry, 2009; Munshi, 2004). Literature also provides theoretical and empirical evidences about the benefits of risk sharing network in rural

areas, thus, it could be another motivation for household participation (Carter and Castillo, 2005; Fafchamps and Gubert, 2007; Karlan, 2007). Finally, besides these 'selfish' motives, pure altruism also can induce active community participation (Andreoni, 1990). Overall, participation may also enhance people's abilities to become more sociable and so, act as a building factor of social capital. Extensively, social capital is itself associated with a wide range of outcomes, of which helping people gain influence and power as well as developing solidarity among the society (Coleman, 1988). Thus, it is expected that household participation will not just bring household direct benefits from a specific project, as proposed by CDD but there are also other potentially indirect effects based on different motivations mentioned above.

A number of empirical studies have shown the positive effects of household participation. Labonne and Chase (2009) using data from 12000 households participating in a CDD project in the Philippines, analyze how communities select their proposals and how resources are allocated across villages. The assumption is that richer and better educated individuals, the elites, are more likely to have influence over project selection and that poorer villages are more likely to be selected for funding. For each household, they construct two dummy variables of whether their preferences were heard. The first indicates if the household cited what the village proposed as one of the three most pressing problems facing the village. The second dummy indicates if the household cited what the village proposed as the most pressing problem. They then conduct a probit regression on the probability that the proposal represents the elected village leader's preferences but not the community preferences. They find that the more unequal the village, the more likely the village leader is to override the community preferences during the proposal selection. But, the greater the proportion of villagers engaged in informal collective action activities, the less likely is the elected village leader to override community preferences during proposal selection. Controlling for poverty, more unequal villages are more likely to receive funding.

Using data collected from 45 villages in two different water supply projects in India, Prokopy (2005) evaluates whether participation in general has a relationship with project outcomes, and if higher levels of participation lead to improved outcomes. The lowest level of participation is passive participation as monetary contribution, and a higher level consists of more active forms of participation such as involvement in decision-making and meeting attendance. The results suggest that overall community participation is fundamentally important to project success and that the greater the number of households who have contributed towards capital cost in a village, the more likely households are to be satisfied and to rate the project as effective.

Examining the relationship between participatory variables and beneficiary satisfaction with the work of the water management committees in Kenya, Ananga (2015) finds a significant association. The hypotheses are that community participation will lead to increased beneficiary satisfaction with the work of the water management committees. Households participating in water meetings and relying on community managed water schemes will tend to practice better water handling hygiene than households who do not. There are several participation-related factors which may affect the performance of urban-based community operated water schemes. The different community participation indicators include participation in decision making, informed choice, economic contributions, representation and responsibility. Results from logistic regression models indicate that households who were actively participating tended to be more satisfied with the work of the water management committees.

These positive effects, however, can be undermined by different circumstances that can cause household participation to be ineffective. Limitations of participatory approach often include various forms of 'elite capture' by the non-poor or majority groups that can hinder the participatory process and the benefits of the program will be captured by the commune

authorities. Addressing the issue of elite capture in CDD projects, Beard and Dasgupta (2006) examines the vulnerability of community driven poverty alleviation approach to elite capture. Participants having asymmetrical social positions and different literacy rates, they enter the process form unequal positions of power which renders community governance particularly vulnerable to elite capture. They use the case study of an Urban Poverty Project in Indonesia as an empirical lens. The findings nuances the often assumed relationship between a community's capacity for collective action and elite capture, Not all elites who had power were corrupt, a finding that highlights the important distinction between elite control and elite capture. In fact, according to Mansuri and Rao (2013), the poor tend to benefit less from participatory process due to the fact that decisions typically reflect the preferences of elite groups. Elite capture tends to increase in high-inequality communities and communities which are remote from centers of power, have low literacy or have significant caste, race and gender disparities.

Mansuri and Rao (2013) highlight four general concerns and limitations to participatory approach. The first is that inequality tends to worsen efficiency and equity. Second, the transfer of management responsibilities to beneficiaries may require the creation of local management capacity. In the absence of efforts to create such capacity, investments in infrastructure are largely wasted. The third is that decentralization can be more a mechanism for tightening central control often because it is a holdover from colonial rules designed to extract resources from local communities rather than to benefit them. Finally, in order for communities to benefit from the resources they manage, governments or implementing agencies need to sustain engagement to build local capacity and to monitor outcomes. As Mosse (2001) argues that several participatory projects that had a high level of participation, "*was often a construct of the planning context and concealed the underlying politics of knowledge production and use*". He raises some arguments about project facilitators who shape and direct the participatory processes, the local collusion in planning, manipulation in project selections, etc.

The literature thus highlights that institutional reform to internalize the complexity of beneficiary participation as well taking into account the context can help making participatory development projects much more effective. To have a comprehensive understanding of the effects of the participatory process in our paper, it is thus important to examine how the participation approach is applied in the context of Vietnam.

## 2.2 CDD and participation approach in Vietnam context

In Vietnam, participatory approaches emerged at the beginning of the 1990s when the political climate became more favorable under the "Doi moi" (renovation) policies initiated by the Vietnamese government after the late 1980s (Neef et al., 2013). Participation in Vietnam has become a society concern with the promulgation of the 1998 Decree on Grassroots Democracy by the Vietnamese Government, aiming at enabling citizens to exercise their rights, be informed of government activities that affect them, discuss and contribute to the formulation of certain policies, participate in local development activities and to supervise certain government actions. In addition, the Government also promoted the principles of participatory process in the country's poverty reduction strategy with the Participatory Poverty Assessments (PPAs) initiatives (Culas et al., 2015). With the country's strong effort to adopt the rhetoric on "good governance", Vietnam seems to be providing a favorable environment for the implementation of CDD in its poverty reduction programs.

In the field, the implementation of participatory approach is however, still limited and under the debate of a civil society concept. In 2006, the UNDP run an exhaustive qualitative study to assess the implication of the Decree on Grassroots Democracy and some other PPAs initiatives in Vietnam. According to this report, women participation is limited in decision making at the lower

level of administration unit and village affairs; people in rural area may be too isolated to participate in decision making since they live far from the commune's center. Also, in rural areas, ethnic minorities suffer from social exclusion that can limit their participation to public life, etc. In a study on the impact of local democracy legislation on people's participation in several communes (Oxfam GB & Institute of Economics), it is found that there have been positive impacts on the quality and capacity of the local administrative and political apparatus: increasing information provision and consultation with local communities on plans and decisions, establishment of community-based 'self-management groups', and a positive impact on the more equitable distribution of resources for poverty reduction. The study finds, however, that the participation of women, poor households, and ethnic minority households is often still limited.

Moreover, participation in Vietnam is still dictated by fundamental dimensions of the Vietnamese culture: "the history of Confucianism in Vietnam, combined with ideas related to the role of the family and elders, has reinforced traditional respect for authority" (UNDP, 2006). The history of central planning has made people relying on the role of the officials in decision making. The report also emphasizes on the fact that weak local governance could harm participation: "lack of clarity of legal instruments, insufficient funding and overlapping roles and responsibilities among government institutions and mass organizations remain obstacles to more active engagement of citizens in their political institutions" (UNDP, 2006). In Vietnam, participation is highly associated with 'mobilization', conducted by mass organizations which are under the Vietnam Fatherland Front and belonging to the Communist Party. Indeed, this strong mobilization has leaded to the common 'voluntary participation' in Vietnam, for instance, volunteers for 'donations of solidarity' can be close to 100% (Culas et al., 2015). For CDD approach implementation in Vietnam, the question is also raised on the side of local authorities. As mentioned above, the history of central planning has affected how local authorities work. They are under pressure from the higher authorities and also evaluated by higher level, thus, they need to mobilize their community to participate in order to meet the 'target' or 'quotas'.

In short, Vietnam has widely adopted the CDD principles in its development projects, however, the impacts of this approach has not yet investigated on household's economic and welfare outcomes. Given its specific political and economic context, it is important to examine how CDD approach is implemented in the field and whether it brings any effect on economic and welfare outcomes for participants.

# 3 The data and empirical strategy

## 3.1 Data source

This study uses data from the Baseline Survey and Endline Survey of the Program 135-II (P135-II) in 2007 and 2012, respectively. The Program 135 (P135) is a targeted program for the Extremely Poor Communes in the Ethnic Minority and Mountainous Areas of Vietnam. As a commune targeted program, the main purpose of P135 is to provide local infrastructure in those areas (such village roads, communes roads, irrigation system, schools, etc.) meanwhile as in other household targeted programs such as the Hunger Eradication and Poverty Reduction program (HEPR), the main objective is to provide the poor with access to credit, education, health services, etc. (Nguyen et al., 2017) The first phase of P135 was conducted in the period of 1998 to 2005 and the second phase in the period of 2006 – 2010.

Success of the first phase (1998 – 2005) created new challenge for the following poverty reduction program. Amongst those new challenges, building and applying the principle of commune ownership, a key part of the decentralization approach, was an important basis for implementation of the Program 135 – phase II as the lack of participation in the first phase of the Program resulted in limited impacts on the program outcomes. In this period, commune ownership means that the commune authorities were given autonomy in planning and implementing projects. Therefore, the project would best meet the need of the community. In addition, as ownership also implied direct responsibilities of the community to the projects, we expect that the project would be better monitored (Herrera et al., 2008).

In the second phase of the Program, its objectives were extended to four components, including infrastructure, agricultural production, socio-cultural livelihoods and capacity building. Therefore, community driven approach was promoted in all phases of planning, managing, supervising and operating infrastructure projects in order to improve resource allocation and effective implementation of infrastructure projects. This also implies an objective of improvement in transparency and accountability in project implementation towards poverty reduction sustainability. The Program P135 indeed bears similarities of its principle to community driven development approaches to the extent that it adopts a decentralized management and planning, as well as promotes active participation of local communities in the selection, supervision and management of small-scale infrastructure schemes. High degree of administrative decentralization permits transfer of responsibility to the local political echelons in the form of people's committees and councils, which operate at the different administrative levels (province, district and commune).

To evaluate the effectiveness of this program and to enhance the design of future program, the Committee for Ethnic Minorities with the support of UNDP implemented a Baseline Survey in 2007 (BLS) and an Endline Survey in 2012 (ELS). The surveys covered 400 communes[1]. All communes are among the P135 phase I. Then, one village randomly selected from the list of all villages in each commune, from which 15 households were selected in each village which amounts to almost 6,000 households. The 2012 ELS followed these households and there are 5,668 households covered in the 2012 ELS since other households migrate and cannot be tracked. Data were collected using household and commune questionnaires.

These data sets are the most complete and comprehensive data sets on ethnic minorities and on the poorest communes in Vietnam. Thus, they provide a thorough understanding of ethnic minorities' socioeconomic situations. For the purposes of our study, we focus on the Participation component, in which, we obtain information about all infrastructure projects implemented in these communes over the 5 year period, from 2007 to 2012. Each household is asked about which construction has been done in their commune, then, a series of questions about participation is provided, following the requirements of participatory approach, from project design to the project construction. From this data component of how infrastructure project is implemented over 5 years, we aggregate information on participation at household level. For the purpose of examining the effect of participation at household level, we take into account three types of treatment: i) if household attends projects' selection meeting; ii) if household expressed their opinions during the projects' selection meeting and iii) if household contributes in cash or in kind/labor to the project. Household considered as being treated if they exercise one of the three measure of participation at least once. This information is extracted from the 2012 ELS. After excluding missing values, we have a total of 8269 observations in an

[1] Details of the sampling survey are in Phung et al. (2012)

unbalanced panel[2], with 4013 observations in 2007 and 4701 observations in 2012. Regarding the first measure of participation, in 2007, we have 2047 treated observations and 1746 controls; this number is 2299 and 2177, respectively in 2012. For the second measure of participation, we have 835 treated and 2958 control observations in 2007; 921 vs. 3555 observations in 2012. For the last measure of participation, we have 1193 treated and 2600 control observations in 2007; the corresponding number is 1367 and 3109 in 2012.

## 3.2 Empirical strategy

### 3.2.1 Assessing the effects of household participation

*Measure of household participation*

In poverty reduction programs, participation can be considered as a way to expand the poor's capacities (Narayan, 2000) and empowerment becomes the objective of these projects. However, to assess the participatory component, before going to find a feasible measurement of participation, the question is whether we consider participation as "an end or a means to an end" (Khwaja, 2005). On one hand, if empowerment is seen as an internal part of agent's utility or welfare, evaluation on participation will purely focus on the level of participation and should not care about how improved capacities help people to obtain better economic outcomes. On the other hand, if empowerment is regarded as a means to increase agent's utility or welfare, empowered agent will be able to get higher incomes, education level or health, etc. Khwaja (2005) expresses mathematically the idea as follows:

Empowerment as an ends: $U_i \equiv f(E_i, X_i)$

Empowerment as a mean to an ends: $U_i \equiv g(X_i)$ and $X_i = h(E_i)$

Where U is a utility function of individual I, E is level of empowerment and X is a list of factors affecting her utility while f(.), g(.) and h(.) are functions. The conditions are $\frac{\partial f}{\partial E} > 0$ and $\frac{\partial h}{\partial E} > 0$. When empowerment is regarded as a mean to an end, it is needed to establish a link between participation measurements to economic outcomes in order to find any causal effects. This reflection is complementary and in line with the two dimensions proposed to assess participatory component in poverty reduction programs by Culas et al. (2015). They include intrinsic value (how far is participation really operating in the field?) and instrumental value (does participation make a difference on the economic outcomes?).

Our next question is how to measure community participation, in particular, household participation. The literature on empowerment suggests four types, including: (i) Passive access which refers to a person's presence in any domain at any level; (ii) Active participation which refers to an actor's expression of voice; (iii) Influence which means the capacity to have one's voice taken into account and (iv) Control represents a position of ultimate power for a group or individual. However, in a similar study like us, Prokopy (2005) considers money contribution as the lowest level of participation, and a higher level consists of more active forms of participation such as involvement in decision-making and meeting attendance. In our study, due to the specific context of Vietnam, we use the three main measures for participation collected in the surveys

[22] For an alternative, we run all regressions with a balanced panel and the results are similar. Alternative results are upon requests to the author.

without categorizing into passive and active forms, including household participation in project meetings, expressing opinions in the meetings and contribution to the project.

*Final outcomes and intermediate outcomes*

Regarding outcome measures, our limitation in this study is that we do not have outcome information at project level, thus, we are not able to directly test the effect of household participation in any specific project's outcomes. Instead, we evaluate the effect of participation at household level as 'an instrument to meet an ends'. We first examine the effect of participation on final outcomes which are economic welfares. Income per capita is the main indicator for this measure, given the fact that expenditure data is not collected in this survey. Apart from income, ownership of durable goods and assets remains an important indicator of households' economic well-being. Ownership of durables and assets reflects the extent to which their income is increased. We use log of per capita net income, assets and durables as the final outcomes in our analysis.

In the second step, we explore different possible channels through which the effects of participation could occur. The first set of intermediate outcomes include access to basic infrastructure and public goods, namely access to electricity, clean water and sanitary toilet. Then we examine a second set of intermediate outcomes associated with better communication, thanks to the proper choice of roads, schools, art centers, etc. A road construction in a village, for instance, helps to reduce the time of transportation, improve access to several services and open more economic activities, especially non-farm work and access to credit. Thus, assessing the intermediate outcomes could potentially explain the effects of participation on final outcomes. Based on the data availability, we can examine a set of intermediate outcomes including acquisition of bike, motorbike, mobile phone, access to agricultural information and access to credit. Lastly, we investigate how participation affects households' self-assessment on their difficulties, namely their cash shortage, difficulties in water access, education payment and medical payment.

*Difference in Difference with Fixed Effect model*

We start our estimation by a difference-in-difference (DiD) approach to compare the changes in outcomes from 2007 to 2012 between those who participate and those who do not. The benefit of using a DiD model is that we can get rid of selection bias which must be time-invariant under the identification assumption that in the absence of participation, the treatment and control group would follow the same trend in outcome growth. We improve the estimation by controlling for all time-invariant characteristics, including the unobserved ones, through a fixed effect model. We can then control for the fact that initial conditions of explanatory variables may have an impact on the subsequent changes in performances or decision to participate. The DiD and fixed effect model is:

$$Y_{it} = \beta_1 P_{it} + \beta_2 t + \beta_3 P_{it} t + \beta_4 X_{it} + \pi_i + \varepsilon_{it} \qquad (1)$$

where $Y_{it}$ is the outcome of the household i at time t. $P_{it}$ is a dummy variable of participation, which captures the possible differences between participants and non-participants in the absence of treatment; t is a time dummy, capturing aggregate time factors that could affect outcomes. We also use $X_{it}$, a set of control variables (household and commune characteristics) in order to increase the precision of our estimates regarding household's and commune's time variant characteristics. And $\pi_i$ here illustrates the household fixed effects that help to address the selection bias on the unobserved variables.

This is a DID model in an OLS setting and $\beta_3$, the coefficient of the interaction between participation and time dummy is our coefficient of interest. The OLS estimate $\beta_3$ can be expressed as:

$$\beta_3 = (\bar{y}^p_{2012} - \bar{y}^p_{2007}) - (\bar{y}^c_{2012} - \bar{y}^c_{2007})$$

where $\bar{y}^p_{2012}$ and $\bar{y}^p_{2007}$ are the average outcome of participants in 2012 and 2007, respectively and $\bar{y}^c_{2012}$ and $\bar{y}^c_{2007}$ are the average outcome of non-participants in 2012 and 2007.

The control factors $(X_{it})$ are used in order to narrow the omitted variable bias due to time-variant factors. We add progressively the households' and then communes' characteristics in the estimations. The household characteristics include age, sex of the head, his/her education level (completed grade), occupation (farming or not), ethnicity, language spoken (Vietnamese, ethnic language), and household size. The commune characteristics include population of the commune, a dummy determining if there is a craft village in the commune, frequency of agriculture extension officials, ethnicity (Kinh or non-Kinh) and education of the commune's leader. Regarding the infrastructure of the commune, we add a dummy for each type of project constructed (if any kind of road was constructed; school or cultural house, health centre), the number of infrastructure projects built in the commune (in log) and lastly the score of how far CDD is implemented in these projects (detail of this measure is described in the next section). There are two main reasons for including these commune's control variables. First, by controlling for infrastructure availability in the commune, we expect to mitigate the bias due to the effect of infrastructure benefit which is correlated with both household's participation and their economic outcomes. The second bias is related to the management ability of the commune officials. If local officials manage development projects well, they are likely to be able to motivate people to participate and at the same time, affect the economic growth of their community. Adding CDD score, ethnicity and education level of the commune's leader enables us to disentangle the effect of household's participation and commune's management capacity.

Finding that households participating in the projects are generally richer and often come from Kinh ethnicity, we suspect that self-selection may lead to a heterogeneity of participation effects. In order to control for initial positions of the households, we examine whether there is any heterogeneity effect of participation for the poor and ethnic minority in order to identify the difference in participation impacts between poor and non-poor households, between minority and non-minority households, but also between the households who are in a P135-II commune and those who are not.

The test is implemented by adding a dummy variable, denoted "Poor[33]", indicating if households are categorized as poor in 2007. The following regression is performed:

$$Y_{it} = \beta_0 + \beta_1 P_{it} + \beta_2 Poor + \beta_3 P_{it} Poor +$$

$$\delta_0 t_{it} + \delta_1 t_{it} P_{it} + \delta_2 t_{it} Poor + \delta_3 t_{it} P_{i1} Poor + \gamma X_{it} + \pi_i + \varepsilon_{it} \qquad (2)$$

In this specification, $\delta_3$ illustrates a triple difference:

$$\delta_3 = \left[(\bar{y}^p_{2012} - \bar{y}^p_{2007}) - (\bar{y}^c_{2012} - \bar{y}^c_{2007}) \middle| P_{i1} = 1, Poor = 1\right] - \left[(\bar{y}^p_{2012} - \bar{y}^p_{2007}) - (\bar{y}^c_{2012} - \bar{y}^c_{2007}) \middle| P_{i1} = 1, Poor = 0\right]$$

[33] Poor household is identified based on the commune's list of poor household at the baseline survey. This classification is based on poverty line and the lack of access to some basic infrastructure and social services.

This coefficient means that for a participating household (as compared to non-participating), belonging to the poor category has a positive or negative effect on their outcomes (relatively to participating households who belong to the non-poor group), i.e on their progress over time. Similar regressions are implemented by replacing the "Poor" with "Minority" indicating the household belongs to ethnic minority group, then "P135" indicating the household belongs to a P135-II commune. From 2007 to 2012, 266 communes from the whole samples of 400 communes are under the proverty reduction program-P135-II, thus, these communes are supposed to receive more infrastructure projects and the CDD approach is one of the P135-II's target. We expect to find higher benefit of participation for participants living in P135-II communes.

*Difference in Difference with Matching method*

The DiD estimation with FE is based on the common trend assumption that treated and control groups would have followed the same paths in the absence of treatment. However, this assumption would be violated if there are some initial differences between the treated and control groups that can affect their outcome evolution differently. In order to mitigate this selection bias, we use propensity score matching (PSM) technique to balance the samples. We select participants and non-participants as similar as possible in terms of observed characteristics that might affect both their decision to participate and their outcomes. However, PSM is based on only observable characteristics, which would lead to selection bias if indeed the participating households and the comparison households are different along unobservable characteristics. Thus, we combine PSM and DiD models to overcome this selection bias concern. We use two methods in PSM to rebuild the counterfactuals, including reweighting and propensity score matching. First, we match based on households' characteristics, including age, gender, education, ethnicity and occupation of the household's head, then, household size and poverty status. We further use commune characteristics for matching such as population of commune, land area, infrastructure projects implemented, information about ethnicity and education of commune's leader. A logit model is estimated to calculate the propensity score and we control for initial conditions that might affect households' trajectories. Then, we reweight our observations by the odd-ratios of the score estimated and restrict our sample to those we have common support. In the second method of matching, we use propensity score matching as a robustness check.

### 3.2.2 Assessing the effect of CDD approach at commune level

In the previous part, we focus on the effect of participation from households' perspective. In this section, we aim at providing a more comprehensive picture about how CDD approach is applied in each project in order to provide an aggregate index of CDD implementation at a commune level. Then, we expect to test the effect of CDD applied projects on the average outcome indicators.

How is the CDD approach implemented in the field? CDD, as shown in section 2, relates to both household's and commune's practices. The advantage of our data regarding to CDD is that, we have inclusive information about CDD provided by local households about each infrastructure project. In this section, we construct an index of CDD for each project and then aggregate it at the commune level. This CDD score, is expected to reflect more objectively the The first indicator is consultation meeting, whether a meeting is organized to choose the project or not. Secondly, we choose the indicator of public bidding, if the commune organizes public bidding for the construction. Third, community supervision board is included. The fourth component is that if there is any contribution, in cash or in labor, by local community. The last indicator is public dissemination of financial information of the project. Those indicators are chosen based on the common requirement for CDD projects as mentioned above. Only two

requirements are not included in calculating this score is the creation of an operational project management unit and operation and maintenance plan since this information is more technical and local households are not aware of. Accordingly, these indicators are not asked in the households' questionnaires.

We use Multiple Correspondence Analysis (MCA) to build a synthetic indicator of CDD practices for each project in each year and this proxy for, to what extent; CDD is applied in the local context. We then use this score as an explanatory component of economic growth at the commune level in the estimation. CDD score for each infrastructure project is thus, computed by the weighted sum of its response to each indicator, such that

$$\mathrm{CDD_i} = \sum_{j=1,..,5} \mathrm{D_{ij}} \mathrm{W_j}$$

Where $\mathrm{CDD_i}$ is the CDD score of a project i, Dij is the response of project i to dimension j, and Wj the MCA weighted for the first axis applied to indicator j. Statistically, the highest weights correspond to the participative behaviors highly correlated with other forms of CDD approach but not very common among all project; and inversely for the lowest ones.

To examine the effect of CDD approach with a purpose of comparison with the household's participation section, we use the same outcomes as before but aggregated at commune level. Indeed, in each commune, there are different projects, thus, once we measure the CDD score for each project, we take the average score in each commune, indicating how CDD approach is implemented in a commune. It is noted that both outcomes and CDD score are standardized since they cannot be directly interpreted in levels. Regarding the control variables, we use aggregated variables (average) from household's data such as age, gender of the household's head, education level, household size, farming occupation and if household receives agriculture information. We also use additional commune controls taken from the commune data, including the commune size, main ethnic group in the commune (categorized into 8 groups), land areas. Besides, to control for agriculture extension promotion during the 5 year period, we also add information on the frequency of agriculture official visiting local households and whether there is a craft village in the commune which is a proxy for job opportunity. Lastly, we include information about ethnicity of the commune' chair or vice chair (a dummy if he or she is Kinh or not) and education level of this highest commune head to partly control for differences in local authority's ability in managing projects. It is noticeable that even Kinh as the majority ethnic group in a commune accounts for around 18 percent of all communes; Kinh being a commune's top three leaders make up more than 25 percent of the communes. The highest positions in these communes are also dominated by man with 94 percent over all three top positions. Other than ethnicity, gender and education level, there is no more information about the officials in these communes. Once we limit our estimation to commune level estimation, the number of communes with all information mentioned above in both 2007 and 2012 is 705 in total, thus, our regressions only use district fixed effect rather than commune fixed effect, including a time trend.

In order to separate the effect of CDD variation from the variation of different infrastructure projects built in each commune, we also add a dummy for the main four types of project: roads, schools and culture houses, medical projects and other types. Lastly we include a dummy for P135-II commune and an interaction between CDD score and treatment communes, since CDD is aimed to be strongly promoted in those P135-II communes. Our purpose of estimation at commune level is to examine the association between CDD implementation and communes' economic and social outcomes, we do not aim at claiming for causality.

# 4 Descriptive statistics

## 4.1 An overview of household participation in infrastructure projects

In this section, we present a descriptive analysis comparing treated and control groups, along three types of participation, namely attending the meetings, expressing opinions in the meeting and contributing to projects. Their differences in characteristics and outcome evolutions are the core of this analysis.

*i. Who are the ones participating in infrastructure projects?*

Results in Table 1 shows that there are some key characteristics associated with household participation, no matter which kind of participation is measured. The participating households are usually headed by male, having higher level of education and they can speak Vietnamese. Meanwhile, ethnicity does not seem to play a role in determining household participation, we find that the proportion of ethnic minority in treated and control groups is somewhat equivalent. Except in the case of raising voice; ethnic minority seems to less be willing to raise voice to express their demand as there is a significant difference between the treated and control group. Similarly, we do not find that the poor participate less than non-poor in terms of attending selection meetings and contributing to the project in cash and or labor. They are less likely to express their opinions than the non-poor households. The proportion of households who live in their commune for a longer period is higher among the participants. Regarding their job, it appears that households who are mainly engaged in agricultural activities participate less in attending selection meetings and raising voices in the meetings.

In the context of Vietnam, participation in development project is questioned by the fact that households might be 'mobilized' or 'encouraged' by the local officials to participate due to the communes or villages' strength of management. Thus, we also illustrate the differences between households who participate and who do not, by communes' characteristics. Characteristics of commune leaders' in terms of ethnicity and education level are not strongly associated with household participation. Households who participate often live in smaller size communes where there are higher number of infrastructure projects and higher frequency of agricultural extension official visiting or contacting local farmers. Lastly, the region seems to play a role in household participation. In the North and in the South, the proportion of households going to the meetings and raising voices is lower. In the Central, people seem to participate more, except for the last indicator of contributing to the infrastructure projects. Evidence from this descriptive statistics indicates that there are indeed some differences in baseline characteristics between the treated and control groups. Households equipped with better human capital like language, education, are more likely to participate.

**Table 1 Baseline characteristics of households (2007) by different measures of participation**

| | Attending meetings | | | Expressing opinions | | | Contributing to project | | |
|---|---|---|---|---|---|---|---|---|---|
| | Treated | Control | P value | Treated | Control | P value | Treated | Control | P value |
| Observations | 2047 | 1746 | | 835 | 2958 | | 1193 | 2600 | |
| *Household characteristics* | | | | | | | | | |
| Age | 40.41 | 39.31 | 0.001 | 42.32 | 39.24 | 0.00 | 40.13 | 39.76 | 0.289 |
| Male headed | 0.92 | 0.88 | 0.00 | 0.92 | 0.89 | 0.008 | 0.92 | 0.89 | 0.004 |
| School years | 4.93 | 3.73 | 0.00 | 5.56 | 4.03 | 0.00 | 4.92 | 4.10 | 0.00 |
| Ethnic minority | 0.80 | 0.80 | 0.857 | 0.77 | 0.81 | 0.021 | 0.82 | 0.79 | 0.103 |
| Household size | 4.98 | 5.01 | 0.525 | 5.02 | 4.99 | 0.636 | 4.98 | 5.00 | 0.806 |
| Poor household | 0.44 | 0.46 | 0.253 | 0.41 | 0.46 | 0.005 | 0.43 | 0.46 | 0.05 |
| Farmer | 0.83 | 0.86 | 0.004 | 0.79 | 0.85 | 0.00 | 0.84 | 0.84 | 0.856 |
| Speaking Vietnamese | 0.36 | 0.25 | 0.00 | 0.43 | 0.28 | 0.00 | 0.34 | 0.29 | 0.008 |
| Years living in commune | 11.51 | 11.36 | 0.60 | 12.27 | 11.23 | 0.00 | 12.73 | 10.88 | 0.00 |
| *Commune characteristics* | | | | | | | | | |
| Commune leader is Kinh | 0.26 | 0.26 | 0.752 | 0.28 | 0.25 | 0.164 | 0.23 | 0.27 | 0.003 |
| School years of leader | 10.33 | 10.13 | 0.001 | 10.32 | 10.21 | 0.119 | 10.31 | 10.19 | 0.08 |
| Number of households in commune | 715.82 | 804.05 | 0.00 | 700.53 | 773.68 | 0.001 | 706.69 | 781.36 | 0.00 |
| Number of projects | 3.2 | 3.04 | 0.01 | 3.36 | 3.06 | 0.00 | 3.2 | 3.01 | 0.12 |
| Agriculture officials visits per year | 19.61 | 17.50 | 0.07 | 21.5 | 17.8 | 0.01 | 22.3 | 16.98 | 0.00 |
| Region: North | 0.58 | 0.71 | 0.00 | 0.54 | 0.67 | 0.00 | 0.69 | 0.62 | 0.00 |
| Region: Central | 0.36 | 0.20 | 0.00 | 0.40 | 0.25 | 0.00 | 0.28 | 0.28 | 0.965 |
| Region: South | 0.05 | 0.09 | 0.00 | 0.06 | 0.08 | 0.044 | 0.02 | 0.09 | 0.00 |

Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculations

*ii. Outcome evolution*

Poverty reduction programs aim at providing basic infrastructure in order to give local community access to better economic opportunities and increase their ability of managing economic activities. The final purpose is to increase their economic welfare, thus, poverty reduces. Therefore, this descriptive analysis on outcome evolution starts with the final outcomes, measured by log of household's income per capita, durable index and asset index (detail of these measurements is presented in Appendix 1). Initially, under the three kinds of participation, treated households are generally better off in terms of income per capita, durable and asset acquisition (Table 2). From 2007 to 2012, the economic disparities between the two groups still hold. For households who attended meetings, they experienced a significantly larger increase in income per capita and durable ownership than households who do not go to the meeting. Same pattern in income evolution is also observed for the other two types of participation. Both groups, regardless of participation types, experience a slight reduction in their asset index. In

short, participating households improve their economic performance better than non-participating households.

In order to reach the final outcomes, treated households are expected to experience different changes in intermediate outcomes and/or other welfare indicators. These outcomes are directly associated with the projects that they participate in and the intermediate outcomes (if any) could be the channels explaining partly better performance of participating households compared to non-participating households. Details of the findings are in Appendix 2. First, participating households in terms of meeting attendants and voice expression, had significantly more access to basic infrastructure such as electricity, piped or well water or sanitary toilets. The gap reduces after 5 years for only water access indicator. For instance, in 2012, 85 percent of household going to project selection meeting have access to electricity but the figure is only 77 percent for the non-participant group. A different path is, however, observed for the last treatment. Households who contribute to the project in cash or labor do not have better access to electricity, water or sanitary toilet, etc. and there is no remarkable improvement observed for them after 5 years.

Participating households also have better means of transportation and communication such as motorbike, bike and mobile phone. There is a remarkable change in households' having mobile phone. Only 7 percent of household in treated group had a mobile phone in 2007 but 73 percent has one in 2012, the figure is 5.6 percent and 65 percent, respectively for control group (the first measure of participation). Additionally, participating households have better access to agricultural information and access to credit, compared to non-participating households. The gap between two groups is larger in the first two measures of participation than in the case of household's contribution to infrastructure project.

Lastly, when households are asked to assess their difficulties and shortages in terms of cash, water access, payment for children's education and payment for medical and health care, participating households seem to face less problems. However, with self-assessment indicators, the differences between two groups are relatively small and not always significant.

**Table 2 Economic outcomes by three measures of household participation**

| | Attending project selection meetings | | | | | | Expressing opinions in the meetings | | | | | | Contributing to infrastructure projects | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2007 | | | 2012 | | | 2007 | | | 2012 | | | 2007 | | | 2012 | | |
| Variable | Treated | Control | P-value | Treated | Control | P-value | Treated | Control | P-value | Treated | Control | P-value | Treated | Control | P-value | Treated | Control | P-value |
| Log of net income per capita | 7.83 | 7.72 | 0.00 | 8.37 | 8.18 | 0.00 | 7.91 | 7.74 | 0.00 | 8.50 | 8.22 | 0.00 | 7.86 | 7.73 | 0.00 | 8.40 | 8.22 | 0.00 |
| | (0.73) | (0.68) | | (0.89) | (0.87) | | (0.72) | (0.70) | | (0.87) | (0.88) | | (0.68) | (0.71) | | 0.84 | 0.90 | |
| Log of durable index | 1.61 | 1.40 | 0.00 | 1.79 | 1.54 | 0.00 | 1.72 | 1.46 | 0.00 | 1.85 | 1.62 | 0.00 | 1.56 | 1.49 | 0.00 | 1.72 | (1.64) | 0.00 |
| | (0.67) | (0.68) | | (0.58) | (0.68) | | (0.67) | (0.68) | | (0.53) | (0.66) | | (0.69) | (0.68) | | (0.60) | (0.66) | |
| Log of asset index | 1.01 | 1.03 | 0.35 | 0.99 | 0.99 | 0.83 | 1.11 | 0.99 | 0.00 | 1.07 | 0.98 | 0.00 | 1.08 | 0.99 | 0.00 | 1.08 | 0.96 | 0.00 |
| | (0.62) | (0.59) | | (0.59) | (0.60) | | (0.63) | (0.59) | | (0.58) | (0.60) | | (0.60) | (0.60) | | (0.59) | (0.60) | |

Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculations

## 4.2 Implementation of Community-Driven Development approach at commune level

Descriptive statistics on the standardized scores by year and types of infrastructure (mean of household's assessment for each project in a commune) are provided in Table 3[4]. This CDD standardized score can be interpreted in relative levels. Units with a higher score display a mix of more frequent practices of CDD. It is seen that the CDD score is higher in 2012 and among village's road and village's art center projects. These types of infrastructures are implemented in the village, thus, it's common that they better follow CDD requirements in order to meet the local's needs. On the households' side, these projects also bring direct benefits for them, thus, it would be the case that they participate and remember more about the implementation of these infrastructure rather than project constructed at higher level.

**Table 3 CDD score standardized by year and infrastructure types**

| | N=1 051 | N=2 394 |
|---|---|---|
| Construction /Year | 2007 | 2012 |
| Commune's traffic road | -0.205 | 0.012 |
| Village's traffic road | 0.266 | 0.348 |
| Irrigation | 0.051 | 0.124 |
| Market | -0.422 | -0.476 |
| School | -0.173 | 0.049 |
| Village art center | 0.386 | 0.390 |
| Medical center | -0.189 | -0.290 |
| Clean water | -0.197 | -0.050 |
| Electricity line | -0.324 | -0.317 |
| Local Committee office | -0.213 | -0.411 |
| Total | -0.049 | -0.001 |

Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation

# 5 Results

## 5.1 Effects of household participation on final outcomes

*Effect of household participation by attending project selection meetings*

Result from DID-FE estimations are presented in Table 4A, 4B and 4C. For the first kind of treatment - '*attending projects' selection meeting*', we find a positive effect on household's economic performance with or without controls for households' and communes' characteristics. However, once we add controls for number of infrastructure projects implemented and CDD score (standardized), the positive effect of household participation is no longer significant. Thus, it seems that household participation is highly affected by how commune officials manage their infrastructure projects, CDD implementation, in particular.

[4] Details of MCA statistics are in Appendix 3

We do not find any effect on durables and assets index. Since we do not have information on the values of the assets, the index measure could not effectively reflect the changes in income or economic welfare of households. When we control for household's characteristics (results are not reported in the table), it is seen that households engaged in agricultural activities and households having a big family have disadvantage in economic growth. Meanwhile, household headed by male seem to perform better than household headed by woman.

*Effect of household participation by expressing their opinion in the meetings*

As seen in Table 4B, there is a consistent significant effect of household participation on log of income per capita growth. The effect ranges between 9.8 percent and 12 percent. We also do not find any effect on log of durables or asset index. Regarding households' and communes' characteristic controls, we find similar effects on household's income increase as in the previous model. It is noticeable that when we add the CDD score, the magnitude of participation effect reduces slightly but the coefficient is still significant at 1 percent level.

We then test the heterogeneity of participation effect for different group, following equation (2). The three interactions of participation with poor dummy, ethnic minority dummy and P135-II commune dummy do not show any significant effect. Different groups, therefore, benefit similarly from participation. However, as the majority of households in our sample are from ethnic minority groups, the positive result of participation effect is a good signal for ethnic minorities. Number of studies on ethnic minorities in Vietnam has shown that lack of confidence and being 'hesitant' to leave their village and participate in market place (World Bank, 2009) has been a barrier for them to increase voice and self-determination, as the result, preventing them from achieving economic progress. Thus, this result shows the importance of raising voice. Participation by expressing ones' opinion would help people to express their idea, the infrastructure that they want and they can benefit from. Regarding asset and durables ownership, we do not find any significant effect of participation.

*Effect of household participation by contributing to infrastructure projects*

Regarding our last measure of participation - '*contributing to infrastructure projec*t' in cash or in kind/labor, the significant effect on log of net income per capita only holds before we include communes' controls. Similarly, there is a positive effect of participation on growth of assets, however, it is no longer significant when we control for CDD implementation at commune level. These findings again confirm the highly positive relationship between household participation and commune's implementation of CDD approach.

Among three measures of household participation, only the most active form has a significant effect. There could be two potential explanations. The first one is that different forms of participation would induce different effects. Passive participation might not bring any benefit for the participants. Second, the significant role of CDD at commune level makes us question the meaning of 'voluntary participation' or 'voluntary contribution'. A concern has been mentioned at the beginning of our study is that the high number of households attending meetings or contributing to infrastructure project might be affected by the local authorities. If the local officials need to attain some objectives of 'participation', they would have put pressure on local communities to attend the meetings as well as contribute to the projects in order to meet the communes' target.

**Table 4A Effect of participation on final outcomes (Attending projects' selection meetings)**

| DiD_FE | (1) Log income pc | (2) Log income pc | (3) Log income pc | (4) Log income pc | (5) Log income pc | (6) Log income pc | (7) Log income pc | (8) Log durable index | (9) Log asset index |
|---|---|---|---|---|---|---|---|---|---|
| Controls | | | | | | | | | |
| Time | 0.517*** | 0.549*** | 0.579*** | 0.597*** | 0.628*** | 0.539*** | 0.522*** | 0.126*** | -0.000 |
| | (0.030) | (0.032) | (0.036) | (0.036) | (0.063) | (0.042) | (0.051) | (0.028) | (0.027) |
| Participation | 0.089*** | 0.090*** | 0.054* | 0.031 | 0.023 | 0.076* | 0.057 | 0.022 | -0.001 |
| | (0.027) | (0.028) | (0.030) | (0.031) | (0.077) | (0.042) | (0.054) | (0.021) | (0.022) |
| Minority x t | | | | | -0.039 | | | | |
| | | | | | (0.065) | | | | |
| Minority x Participation | | | | | 0.011 | | | | |
| | | | | | (0.083) | | | | |
| Poor x t | | | | | | 0.122*** | | | |
| | | | | | | (0.045) | | | |
| Poor x Participation | | | | | | -0.090 | | | |
| | | | | | | (0.060) | | | |
| P135-II commune x t | | | | | | | 0.106** | | |
| | | | | | | | (0.050) | | |
| P135-II x Participation | | | | | | | -0.026 | | |
| | | | | | | | (0.065) | | |
| Household controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commune controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *Infrastructure controls* | | | | | | | | | |
| If road constructed | | -0.054* | 0.032 | 0.027 | 0.029 | 0.036 | 0.032 | -0.105*** | 0.021 |
| | | (0.030) | (0.044) | (0.044) | (0.044) | (0.044) | (0.043) | (0.030) | (0.030) |
| If school constructed | | -0.022 | 0.022 | 0.016 | 0.014 | 0.018 | 0.020 | -0.021 | 0.080*** |
| | | (0.025) | (0.030) | (0.030) | (0.030) | (0.030) | (0.030) | (0.022) | (0.022) |
| If medical center constructed | | -0.025 | 0.002 | -0.010 | -0.009 | -0.009 | -0.011 | 0.080*** | 0.042* |
| | | (0.026) | (0.031) | (0.031) | (0.031) | (0.031) | (0.031) | (0.021) | (0.022) |
| Number of projects (log) | | | -0.052* | -0.049 | -0.050* | -0.053* | -0.051* | 0.026 | -0.042* |
| | | | (0.030) | (0.030) | (0.030) | (0.030) | (0.030) | (0.021) | (0.021) |
| CDD score (standardized) | | | | 0.044*** | 0.043*** | 0.043*** | 0.040*** | 0.011 | 0.029*** |
| | | | | (0.013) | (0.013) | (0.013) | (0.013) | (0.009) | (0.009) |
| Constant | 7.661*** | 7.724*** | 7.507*** | 7.543*** | 7.522*** | 7.567*** | 7.599*** | 0.555** | 0.110 |
| | (0.283) | (0.286) | (0.305) | (0.306) | (0.309) | (0.306) | (0.305) | (0.229) | (0.246) |
| Observations | 8922 | 8922 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 |
| R-squared | 0.264 | 0.265 | 0.276 | 0.278 | 0.279 | 0.280 | 0.280 | 0.109 | 0.036 |
| Number of id | 4596 | 4596 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

**Table 4B Effect of participation on final outcomes (Expressing opinions)**

| DiD_FE<br>Controls | (1)<br>Log income pc | (2)<br>Log income pc | (3)<br>Log income pc | (4)<br>Log income pc | (5)<br>Log income pc | (6)<br>Log income pc | (7)<br>Log income pc | (8)<br>Log durable index | (9)<br>Log asset index |
|---|---|---|---|---|---|---|---|---|---|
| Time | 0.543*** | 0.568*** | 0.580*** | 0.591*** | 0.636*** | 0.555*** | 0.527*** | 0.144*** | -0.000 |
| | (0.024) | (0.030) | (0.033) | (0.033) | (0.052) | (0.037) | (0.042) | (0.025) | (0.026) |
| Participation | 0.098*** | 0.122*** | 0.123*** | 0.107*** | 0.022 | 0.107** | 0.121** | -0.037 | 0.000 |
| | (0.032) | (0.032) | (0.034) | (0.034) | (0.080) | (0.044) | (0.057) | (0.023) | (0.027) |
| Minority x t | | | | | -0.056 | | | | |
| | | | | | (0.050) | | | | |
| Minority x Participation | | | | | 0.106 | | | | |
| | | | | | (0.088) | | | | |
| Poor x t | | | | | | 0.074** | | | |
| | | | | | | (0.034) | | | |
| Poor x Participation | | | | | | 0.012 | | | |
| | | | | | | (0.068) | | | |
| P135-II commune x t | | | | | | | 0.096** | | |
| | | | | | | | (0.037) | | |
| P135-II x Participation | | | | | | | -0.014 | | |
| | | | | | | | (0.070) | | |
| Household controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commune controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *Infrastructure controls* | | | | | | | | | |
| If road constructed | | -0.054* | 0.032 | 0.028 | 0.030 | 0.034 | 0.033 | -0.107*** | 0.021 |
| | | (0.030) | (0.043) | (0.043) | (0.043) | (0.043) | (0.043) | (0.030) | (0.031) |
| If school constructed | | -0.021 | 0.019 | 0.014 | 0.013 | 0.015 | 0.017 | -0.021 | 0.080*** |
| | | (0.025) | (0.030) | (0.030) | (0.030) | (0.030) | (0.030) | (0.022) | (0.022) |
| If medical center constructed | | -0.023 | -0.000 | -0.012 | -0.011 | -0.012 | -0.013 | 0.082*** | 0.042* |
| | | (0.026) | (0.031) | (0.031) | (0.031) | (0.031) | (0.031) | (0.021) | (0.021) |
| Number of projects (log) | | | -0.049* | -0.048 | -0.049 | -0.049* | -0.049* | 0.028 | -0.042* |
| | | | (0.030) | (0.030) | (0.030) | (0.030) | (0.030) | (0.021) | (0.021) |
| CDD score (standardized) | | | | 0.041*** | 0.041*** | 0.040*** | 0.038*** | 0.015* | 0.029*** |
| | | | | (0.013) | (0.013) | (0.013) | (0.013) | (0.009) | (0.009) |
| Constant | 7.503*** | 7.688*** | 7.466*** | 7.500*** | 7.487*** | 7.522*** | 7.555*** | 0.588** | 0.109 |
| | (0.261) | (0.287) | (0.306) | (0.306) | (0.309) | (0.306) | (0.305) | (0.228) | (0.247) |
| Observations | 8922 | 8922 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 |
| R-squared | 0.260 | 0.266 | 0.278 | 0.280 | 0.280 | 0.281 | 0.282 | 0.109 | 0.036 |
| Number of id | 4596 | 4596 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

**Table 4C Effect of participation on final outcomes (Contributing to infrastructure projects)**

| DiD_FE | (1) Log income pc | (2) Log income pc | (3) Log income pc | (4) Log durable index | (5) Log durable index | (6) Log durable index | (7) Log asset index | (8) Log asset index | (9) Log asset index |
|---|---|---|---|---|---|---|---|---|---|
| Controls | | | | | | | | | |
| Time | 0.581*** | 0.600*** | 0.614*** | 0.136*** | 0.132*** | 0.135*** | -0.048** | -0.016 | -0.009 |
| | (0.030) | (0.034) | (0.034) | (0.022) | (0.026) | (0.026) | (0.023) | (0.026) | (0.026) |
| Participation | 0.041 | 0.014 | -0.004 | 0.010 | 0.008 | 0.004 | 0.047** | 0.039* | 0.028 |
| | (0.028) | (0.031) | (0.031) | (0.020) | (0.022) | (0.022) | (0.021) | (0.023) | (0.023) |
| Household controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Commune controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *Infrastructure controls* | | | | | | | | | |
| If road constructed | -0.059** | 0.030 | 0.025 | -0.068*** | -0.106*** | -0.106*** | -0.036* | 0.023 | 0.022 |
| | (0.030) | (0.043) | (0.043) | (0.020) | (0.030) | (0.030) | (0.021) | (0.031) | (0.030) |
| If school constructed | -0.020 | 0.022 | 0.016 | 0.007 | -0.019 | -0.021 | 0.046*** | 0.084*** | 0.080*** |
| | (0.025) | (0.030) | (0.030) | (0.017) | (0.022) | (0.022) | (0.017) | (0.022) | (0.022) |
| If medical center constructed | -0.019 | 0.005 | -0.008 | 0.091*** | 0.085*** | 0.081*** | 0.039** | 0.048** | 0.040* |
| | (0.026) | (0.031) | (0.031) | (0.017) | (0.021) | (0.021) | (0.018) | (0.021) | (0.022) |
| Number of projects (log) | | -0.049 | -0.047 | | 0.027 | 0.028 | | -0.044** | -0.042** |
| | | (0.030) | (0.030) | | (0.021) | (0.021) | | (0.021) | (0.021) |
| CDD score (standardized) | | | 0.047*** | | | 0.013 | | | 0.027*** |
| | | | (0.013) | | | (0.009) | | | (0.009) |
| Constant | 7.744*** | 7.526*** | 7.558*** | 0.383* | 0.555** | 0.565** | 0.043 | 0.083 | 0.103 |
| | (0.286) | (0.305) | (0.306) | (0.202) | (0.228) | (0.229) | (0.216) | (0.247) | (0.246) |
| Observations | 8922 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 | 8269 |
| R-squared | 0.264 | 0.276 | 0.278 | 0.107 | 0.108 | 0.109 | 0.034 | 0.034 | 0.036 |
| Number of id | 4596 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 | 4573 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

## 5.2 Effects of household participation on intermediate outcomes

We find a consistent effect of active participation (through expressing opinions) on income growth but the remaining question is that what are the channels bridging household participation in infrastructure projects and their final outcomes? And in the case of no effect on final outcomes, is there any intermediate outcome effect? To test different possible channels for the effect of household participation, we run the same estimation model of DiD with FE. Our intermediate outcomes are dummy variables. Using linear probability model for these binary variables is often subject to the drawback that predicted probabilities can be less than 0 or higher than 1. However, this drawback is not present in the case of DiD estimation.

Tables 5A, 5B and 5C present the effects of household participation (across three measures) on intermediate outcomes. Our results suggest a strong and significant effect of participation, regardless of different measures of participation, on households' access to sanitary toilet and mobile phone. There is no effect on the probability of household having motorbikes but a negative effect on bikes. This could suggest to some extent the changes in income and the changes in local infrastructure. Better commuted roads could motivate local households to buy

other modes of transport than bikes. The ones, who participate to choose the projects that meet their own demand or contribute to the construction of those projects in cash or in labor, can derive higher benefits. A highlighted result is that there is a heterogeneous effect of participation for ethnic minority and poor households in access to sanitary toilet, agriculture information and also credit access. The initial lower endowments of these disadvantaged groups could be the reason for them to benefit more from participation rather than their counterparts.

Significant effect of participation (all three measures) on access to agricultural information and mobile phone is a signal of better communication in local communities. Even we do not directly test the effect of community participation on project effectiveness; this finding could suggest positive impacts of participation to some extent. Among different types of infrastructure project, local roads and villages' art houses are the ones in which households participate the most. If they did choose the most important ones that meet their own needs, once the projects are finished, participating households are unquestionably the first beneficiaries. They can rapidly take advantages of the road for their transport needs, as well as other demand for communication. They are then better connected, as their consumption of mobile phone and their access to agricultural information increases. Having better access to information is also an indicator of higher level of social capital. This could be a result of social interaction induced from participation in community activities (Coleman, 1988). Attending a meeting or expressing opinion in community can be seen as a way to build social capital and expand social network.

Results from our two models of DiD-ME are shown in Table 6. We find consistent results for most of the cases. These additional specifications help to confirm the overall effect of participation, especially in the positive effect of active participation on income growth and better communication through agriculture information access and mobile consumption.

**Table 5A Effects of participation on intermediate outcomes (Attending projects' selection meetings)**

| DiD_FE | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Controls | Electricity | Toilet | Toilet | Toilet | Water | Motor | Bike | Mobile | Agricul. information | Credit |
| Time | 0,120*** | 0.144*** | 0.216*** | 0.158*** | -0.007 | 0.292*** | -0.107*** | 0.632*** | 0.056** | -0.009 |
| | -0,015 | (0.019) | (0.034) | (0.022) | (0.014) | (0.023) | (0.022) | (0.021) | (0.026) | (0.024) |
| Participation | -0.030** | 0.063*** | -0.004 | 0.038 | -0.011 | -0.006 | -0.078*** | 0.052*** | 0.120*** | 0.027 |
| | (0.013) | (0.016) | (0.041) | (0.023) | (0.013) | (0.019) | (0.019) | (0.017) | (0.022) | (0.020) |
| Minority x t | | | -0.091*** | | | | | | | |
| | | | (0.034) | | | | | | | |
| Minority x Participation | | | 0.085* | | | | | | | |
| | | | (0.045) | | | | | | | |
| Poor x t | | | | -0.029 | | | | | | |
| | | | | (0.021) | | | | | | |
| Poor x Participation | | | | 0.055* | | | | | | |
| | | | | (0.031) | | | | | | |
| Number of projects (log) | -0.002 | 0.037** | 0.035** | 0.038** | 0.023* | -0.005 | 0.051*** | 0.043** | 0.146*** | 0.050** |
| | (0.012) | (0.016) | (0.016) | (0.015) | (0.013) | (0.019) | (0.018) | (0.017) | (0.022) | (0.020) |
| CDD score (standardized) | -0.013** | 0.010 | 0.009 | 0.010 | 0.002 | -0.001 | 0.007 | -0.022*** | 0.008 | 0.029** |
| | (0.005) | (0.007) | (0.007) | (0.007) | (0.005) | (0.008) | (0.008) | (0.008) | (0.010) | (0.009) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Constant | 0.418 | 0.073 | 0.046 | 0.072 | 0.573*** | -0.442** | -0.455** | -0.204 | -0.067 | -0.103 |
| | (0.129) | (0.157) | (0.158) | (0.157) | (0.122) | (0.204) | (0.184) | (0.180) | (0.210) | (0.210) |
| Observations | 8381 | 8381 | 8381 | 8381 | 8381 | 8381 | 8381 | 8381 | 8381 | 8381 |
| R-squared | 0.121 | 0.155 | 0.157 | 0.156 | 0.010 | 0.214 | 0.069 | 0.652 | 0.090 | 0.037 |
| Number of id | 4603 | 4603 | 4603 | 4603 | 4603 | 4603 | 4603 | 4603 | 4603 | 4603 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation

**Table 5B Effects of participation on intermediate outcomes (Expressing opinions)**

| DiD_FE | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Controls | Electricity | Toilet | Toilet | Water | Motor | Bike | Mobile | Agricul. information | Agricul. information | Credit |
| Time | 0.112*** | 0.159*** | 0.208*** | -0.009 | 0.296*** | -0.124*** | 0.648*** | 0.096*** | 0.083*** | 0.003 |
| | (0.014) | (0.017) | (0.029) | (0.012) | (0.022) | (0.021) | (0.020) | (0.025) | (0.027) | (0.022) |
| Participation | -0.034** | 0.080*** | 0.011 | -0.017 | -0.035 | -0.104*** | 0.045** | 0.090*** | 0.056 | 0.006 |
| | (0.015) | (0.021) | (0.046) | (0.014) | (0.022) | (0.023) | (0.020) | (0.026) | (0.034) | (0.023) |
| Minority x t | | | -0.062** | | | | | | | |
| | | | (0.026) | | | | | | | |
| Minority x Participation | | | 0.087* | | | | | | | |
| | | | (0.051) | | | | | | | |
| Poor x t | | | | | | | | | 0.026 | |
| | | | | | | | | | (0.025) | |
| Poor x Participation | | | | | | | | | 0.088* | |
| | | | | | | | | | (0.051) | |
| Constant | 0.424*** | 0.057 | 0.033 | 0.577*** | -0.424** | -0.425** | -0.209 | -0.064 | -0.047 | -0.095 |
| | (0.130) | (0.158) | (0.158) | (0.122) | (0.204) | (0.184) | (0.180) | (0.209) | (0.209) | (0.211) |
| Observations | 8381 | 8381 | 8381 | 8381 | 8022 | 8022 | 8022 | 8381 | 8381 | 8381 |
| R-squared | 0.121 | 0.156 | 0.157 | 0.010 | 0.214 | 0.070 | 0.651 | 0.085 | 0.087 | 0.037 |
| Number of id | 4603 | 4603 | 4603 | 4603 | 4574 | 4574 | 4574 | 4603 | 4603 | 4603 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation

**Table 5C Effects of participation on intermediate outcomes (Contributing to infrastructure projects)**

| DiD_FE | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Controls | Electricity | Toilet | Water | Motor | Bike | Mobile | Agricul. information | Credit | Credit |
| Time | 0.110*** | 0.165*** | -0.010 | 0.285*** | -0.134*** | 0.647*** | 0.090*** | -0.008 | -0.043 |
| | (0.014) | (0.018) | (0.013) | (0.022) | (0.021) | (0.020) | (0.025) | (0.022) | (0.031) |
| Participation | -0.016 | 0.033* | -0.010 | 0.011 | -0.039* | 0.036** | 0.083*** | 0.044** | -0.030 |
| | (0.014) | (0.018) | (0.014) | (0.020) | (0.020) | (0.018) | (0.023) | (0.021) | (0.044) |
| Minority x t | | | | | | | | | 0.047 |
| | | | | | | | | | (0.029) |
| Minority x Participation | | | | | | | | | 0.089* |
| | | | | | | | | | (0.050) |
| Constant | 0.409*** | 0.094 | 0.570*** | -0.448** | -0.478*** | -0.191 | -0.030 | -0.099 | -0.046 |
| | (0.130) | (0.158) | (0.122) | (0.204) | (0.184) | (0.180) | (0.210) | (0.210) | (0.211) |
| Observations | 8381 | 8381 | 8381 | 8022 | 8022 | 8022 | 8381 | 8381 | 8381 |
| R-squared | 0.121 | 0.153 | 0.010 | 0.214 | 0.065 | 0.651 | 0.085 | 0.038 | 0.041 |
| Number of id | 4603 | 4603 | 4603 | 4574 | 4574 | 4574 | 4603 | 4603 | 4603 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation

**Table 6 Effect of participation on final and intermediate outcomes (comparison with matching methods)**

| | (1) Log of income pc | (2) Log of durable | (3) Log of asset | (4) Electricity | (5) Toilet | (6) Water | (7) Motor | (8) Bike | (9) Mobile | (10) Agriculture information | (11) Credit |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Attending meetings* | | | | | | | | | | | |
| DiD-FE | 0.031 | 0.022 | -0.001 | 0.110*** | 0.165*** | -0.010 | 0.285*** | -0.134*** | 0.647*** | 0.090*** | -0.008 |
| | (0.031) | (0.021) | (0.022) | (0.014) | (0.018) | (0.013) | (0.022) | (0.021) | (0.020) | (0.025) | (0.022) |
| DiD-ME | | | | | | | | | | | |
| (i) Reweighting | 0.0367 | 0.0286 | 0.0170 | -0.0193 | 0.042** | -0.004 | -0.0120 | -0.06*** | 0.038** | 0.118*** | 0.030 |
| | (0.0329) | (0.0208) | (0.0233) | (0.0130) | (0.018) | (0.013) | (0.0195) | (0.020) | (0.018) | (0.023) | (0.021) |
| (ii) PS matching | 0.046* | -0.015 | 0.001 | -0.028 | 0.048** | -0.007 | -0.029 | -0.022 | 0.037** | 0.145*** | 0.014 |
| | (0.024) | 0.025 | (0.028) | (0.019) | (0.017) | (0.015) | (0.024) | (0.024) | (0.018) | (0.027) | (0.024) |
| *Expressing opinions* | | | | | | | | | | | |
| DiD-FE | 0.107*** | -0.037 | 0.000 | -0.034** | 0.080*** | -0.017 | -0.035 | -0.104*** | 0.045** | 0.090*** | 0.006 |
| | (0.034) | (0.023) | (0.027) | (0.015) | (0.021) | (0.014) | (0.022) | (0.023) | (0.020) | (0.026) | (0.023) |
| DiD-ME | | | | | | | | | | | |
| (i) Reweighting | 0.109*** | -0.026 | 0.019 | -0.022 | 0.067*** | -0.016 | -0.033 | -0.086*** | 0.031 | 0.080*** | 0.008 |
| | (0.035) | (0.023) | (0.027) | (0.015) | (0.021) | (0.015) | (0.022) | (0.024) | (0.020) | (0.027) | (0.024) |
| (ii) PS matching | 0.117** | -0.032 | -0.012 | -0.065** | 0.048** | -0.010 | 0.002 | -0.106*** | 0.066*** | 0.083** | 0.023 |
| | (0.049) | (0.031) | (0.033) | (0.032) | (0.023) | (0.015) | (0.034) | 0.026 | (0.022) | 0.029 | (0.039) |
| *Contributing to the projects* | | | | | | | | | | | |
| DiD-FE | -0.004 | 0.004 | 0.039* | -0.016 | 0.033* | -0.010 | 0.011 | -0.039* | 0.036** | 0.083*** | 0.044** |
| | (0.031) | (0.022) | (0.023) | (0.014) | (0.018) | (0.014) | (0.020) | (0.020) | (0.018) | (0.023) | (0.021) |
| DiD-ME | | | | | | | | | | | |
| (i) Reweighting | -0.005 | 0.007 | 0.043* | -0.006 | 0.018 | -0.006 | 0.007 | -0.008 | 0.020 | 0.068*** | 0.036 |
| | (0.032) | (0.022) | (0.024) | (0.014) | (0.019) | (0.015) | (0.021) | (0.021) | (0.018) | (0.024) | (0.022) |
| (ii) PS matching | -0.005 | 0.021 | 0.058* | -0.017 | 0.024 | -0.022 | 0.004 | -0.003 | 0.034 | 0.089*** | 0.047** |
| | (0.057) | (0.038) | (0.033) | (0.032) | (0.023) | (0.018) | (0.026) | (0.023) | (0.023) | (0.024) | (0.021) |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

*Self-assessment on difficulties*

We examine the effect of participation on households' self-assessment. The results in Table 7A, 7B and 7C show that treated households are less likely to declare difficulties. Households who contribute to infrastructure project have lower probability of having cash shortage or difficulties in paying for their children education as well as medical and health care fees. Participation in terms of attending meetings also decreases the probability of having water shortage. These self-assessment results corroborate the idea that household participation in infrastructure projects to some extent has positive effect on their welfare and living conditions.

Household participation measured in our study is limited in the scope of infrastructure projects. An argument could rely on the fact that a construction like village's road should benefit all citizens living in that village/commune. However, not everyone could benefit as effectively as the others. After controlling for the availability of infrastructure projects and management abilities of the local authorities, we show that active participation has some given impacts on household's welfare.

**Table 7A Effect of participation on self-assessment on difficulties (Attending projects' selection meetings)**

| | Water shortage | Cash shortage | Education payment | Medical payment |
|---|---|---|---|---|
| DiD-FE | | | | |
| Household controls | -0.042** | 0.010 | -0.054*** | 0.003 |
| | (0.019) | (0.016) | (0.019) | (0.018) |
| Commune and household controls | -0.037* | 0.009 | -0.046** | -0.017 |
| | (0.022) | (0.019) | (0.021) | (0.021) |
| DiD-ME | | | | |
| (i) Reweighting | -0.051** | 0.005 | -0.036 | -0.016 |
| | (0.023) | (0.020) | (0.022) | (0.021) |
| (ii) PS matching | -0.052** | -0.002 | -0.028 | 0.007 |
| | (0.025) | (0.026) | (0.029) | (0.024) |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

**Table 7B Effect of participation on self-assessment on difficulties (Expressing opinions)**

| | Water shortage | Cash shortage | Education payment | Medical payment |
|---|---|---|---|---|
| DiD-FE | | | | |
| Household controls | 0.022 | 0.020 | 0.004 | 0.027 |
| | (0.024) | (0.021) | (0.023) | (0.022) |
| Commune and household controls | 0.036 | 0.007 | -0.006 | 0.016 |
| | (0.026) | (0.023) | (0.025) | (0.024) |
| DiD-ME | | | | |
| (i) Reweighting | 0.028 | 0.005 | 0.011 | 0.020 |
| | (0.026) | (0.024) | (0.026) | (0.025) |
| (ii) PS matching | -0.007 | 0.022 | 0.015 | 0.007 |
| | (0.038) | (0.03) | (0.028) | (0.037) |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

**Table 7C Effect of participation on self-assessment on difficulties (Contributing to infrastructure projects)**

| | Water shortage | Cash shortage | Education payment | Medical payment |
|---|---|---|---|---|
| DiD-FE | | | | |
| Household controls | -0.025 | -0.034* | -0.061*** | -0.040** |
| | (0.021) | (0.018) | (0.020) | (0.020) |
| Commune and household | -0.014 | -0.039** | -0.049** | -0.044** |
| | (0.023) | (0.020) | (0.023) | (0.021) |
| DiD-ME | | | | |
| (i) Reweighting | -0.022 | -0.045** | -0.040* | -0.035 |
| | (0.023) | (0.021) | (0.023) | (0.022) |
| (ii) PS matching | -0.021 | -0.053** | -0.064* | -0.043* |
| | (0.023) | (0.024) | (0.034) | (0.024) |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

## 5.3 Effects of CDD implementation at commune level

Results from DiD-FE and DiD-ME do not only provide evidence about the positive effect of household participation but also suggest the key role of CDD implementation at commune level. In this section, we assess directly the relationship between CDD implementation and commune's outcomes. Results from the OLS estimations are presented in Tables 8. It is seen that there is a positive association between CDD score and economic welfare outcomes such as income per capita and asset index. For instance, a standard deviation in CDD score is associated with a 5.7 percent increase in income per capita (again for purpose interpretation, CDD score and dependent variables are standardized). We did also test the association of CDD score and other welfare indicators such as access to clean toilet, primary school and secondary school net enrolment, but we did not find any significant relationship. In addition, there is no specific effect of P135-II communes, thus, we do not report these results here[*].

It is seen that across all regression, the coefficient of average household head's education level is positive and significant and the coefficient of household size is negative and significant. It confirms again the important role of education in economic growth and also implies the disadvantage of having a big size family, in particularly regarding school enrolment. Even if coefficients of different main ethnic group in a commune vary, commune with Kinh majority never perform worse than non-Kinh. Then, the result is confirmed when we use only 2 ethnic categories: commune with Kinh majority and commune dominated by other ethnic groups than Kinh. In these regressions, we find a significant effect of Kinh majority in commune's average outcome. We also test the effect of CDD score by ethnicity; however, we do not find any significant effect. Finally, we do not find any significant effect of the commune's leader ethnicity and education level on the performance of his/her commune. These suggestive evidences of the positive association between CDD implementation and communes' outcome should be considered with caution. First, it is possible that local households only remember and choose the projects that are close to them and from which they benefit. Thus, the positive effect would partly capture the effect of household's subjective assessment on the project rather than the CDD implementation by communes' officials. Secondly, it is possible that their answer is biased

[*] These results are available upon request to the author.

upwards since they would expect to receive other projects in the future, as discussed in the literature (Platteau and Gaspart, 2003).

**Table 8 Effect of CDD implementation at commune level**

| OLS | (1) Income per capita | (2) Income per capita | (3) Asset index | (4) Asset index | (5) Durable index | (6) Durable index | (7) High school net enrolment | (8) High school net enrolment |
|---|---|---|---|---|---|---|---|---|
| CDD score | 0.057** | 0.055** | 0.092*** | 0.086*** | 0.049 | 0.045 | 0.089* | 0.085* |
| | (0.022) | | (0.032) | (0.033) | (0.030) | (0.030) | (0.047) | (0.047) |
| *Households' characteristics* | | | | | | | | |
| Age of Head | 0.007 | 0.006 | 0.003 | 0.005 | 0.013** | 0.014** | 0.005 | 0.004 |
| | (0.005) | (0.005) | (0.010) | (0.010) | (0.007) | (0.007) | (0.011) | (0.011) |
| Head is male | -0.029 | -0.004 | 0.041 | 0.207 | -0.728** | -0.631* | 0.016 | 0.035 |
| | (0.314) | (0.312) | (0.385) | (0.392) | (0.343) | (0.339) | (0.652) | (0.632) |
| Year education of head | 0.066** | 0.063*** | 0.053** | 0.057** | 0.125*** | 0.128*** | 0.158*** | 0.159*** |
| | (0.026) | (0.023) | (0.027) | (0.026) | (0.021) | (0.019) | (0.039) | (0.036) |
| Household size | -0.064* | -0.058* | 0.062 | 0.084 | -0.110*** | -0.099*** | -0.214*** | -0.201*** |
| | (0.034) | (0.030) | (0.048) | (0.051) | (0.041) | (0.038) | (0.057) | (0.053) |
| Head is farmer | 0.265 | 0.214 | -0.089 | -0.173 | -0.835*** | -0.901*** | -0.071 | -0.124 |
| | (0.381) | (0.374) | (0.232) | (0.225) | (0.223) | (0.216) | (0.400) | (0.385) |
| Received agricul. Information | 0.241 | 0.232 | 0.437*** | 0.449*** | 0.113 | 0.122 | 0.181 | 0.188 |
| | (0.285) | (0.275) | (0.144) | (0.144) | (0.109) | (0.109) | (0.173) | (0.171) |
| Number of households | -0.000 | -0.000 | -0.000** | -0.000* | 0.000 | 0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| *Main ethnic group* | | | | | | | | |
| Kinh | | 0.179** | | 0.135 | | 0.036 | | 0.379** |
| | | (0.090) | | (0.121) | | (0.103) | | (0.192) |
| Tay | -0.053 | | -0.229 | | -0.068 | | -0.476* | |
| | (0.101) | | (0.164) | | (0.107) | | (0.273) | |
| Thai | -0.181 | | 0.221 | | 0.128 | | -0.256 | |
| | (0.160) | | (0.212) | | (0.165) | | (0.278) | |
| Muong | -0.389*** | | -0.220 | | -0.042 | | -0.276 | |
| | (0.136) | | (0.203) | | (0.195) | | (0.301) | |
| Nung | -0.164 | | -0.169 | | -0.180 | | -0.532* | |
| | (0.125) | | (0.259) | | (0.165) | | (0.283) | |
| Hmong | -0.139 | | -0.049 | | -0.027 | | -0.284 | |
| | (0.115) | | (0.256) | | (0.159) | | (0.251) | |
| Dao | -0.015 | | 0.126 | | 0.248* | | -0.063 | |
| | (0.109) | | (0.238) | | (0.145) | | (0.251) | |
| Others | -0.176 | | -0.030 | | -0.012 | | -0.402* | |
| | (0.113) | | (0.184) | | (0.142) | | (0.214) | |
| Land area | 0.000 | 0.000 | -0.000 | -0.000 | -0.001* | -0.001* | -0.001 | -0.001 |
| | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Agriculture official visit | 0.000 | 0.000 | 0.004*** | 0.004*** | 0.002** | 0.002** | 0.002 | 0.002 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.002) |
| Craft village | -0.047 | -0.044 | 0.096 | 0.113 | 0.098 | 0.114* | 0.150 | 0.166 |
| | (0.045) | (0.043) | (0.081) | (0.082) | (0.065) | (0.066) | (0.104) | (0.103) |
| Chairman is Kinh | -0.140 | -0.138 | 0.024 | 0.003 | 0.050 | 0.036 | -0.166 | -0.183 |
| | (0.145) | (0.143) | (0.107) | (0.107) | (0.082) | (0.081) | (0.144) | (0.141) |
| Education years of chairman | 0.020 | 0.020 | -0.022 | -0.020 | -0.020 | -0.019 | 0.016 | 0.017 |
| | (0.029) | (0.029) | (0.021) | (0.021) | (0.016) | (0.016) | (0.029) | (0.028) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| If road constructed | 0.007 | -0.000 | 0.065 | 0.063 | -0.055 | -0.058 | -0.031 | -0.034 |
| | (0.065) | (0.064) | (0.113) | (0.113) | (0.085) | (0.085) | (0.146) | (0.146) |
| If school constructed | -0.158 | -0.158 | 0.015 | 0.006 | 0.022 | 0.016 | 0.044 | 0.036 |
| | (0.192) | (0.188) | (0.075) | (0.075) | (0.062) | (0.061) | (0.099) | (0.099) |
| If medical center constructed | 0.100 | 0.097 | 0.095 | 0.104 | 0.142** | 0.142** | -0.078 | -0.074 |
| | (0.110) | (0.107) | (0.077) | (0.076) | (0.063) | (0.062) | (0.105) | (0.103) |
| Other project constructed | 0.029 | 0.024 | -0.125 | -0.120 | 0.049 | 0.050 | -0.063 | -0.058 |
| | (0.075) | (0.071) | (0.081) | (0.082) | (0.067) | (0.067) | (0.109) | (0.109) |
| Constant | -0.588 | -0.649 | 0.318 | -0.041 | -0.106 | -0.257 | -0.153 | -0.474 |
| | (0.853) | (0.777) | (0.824) | (0.768) | (0.558) | (0.519) | (1.021) | (0.919) |
| Observations | 705 | 705 | 705 | 705 | 705 | 705 | 705 | 705 |
| R-squared | 0.395 | 0.394 | 0.728 | 0.723 | 0.831 | 0.827 | 0.557 | 0.553 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects.
Source: BLS 2007, ELS 2012, CEMA, UNDP, author's calculation.

## 6 Conclusion

In this paper, we assess the implementation of CDD and community participation approach at both household and commune level. Using DiD with Fixed effect and DiD matching estimators, we find that participation only has significant effect on economic outcome when household participate actively in the project, by expressing their opinions in projects' selection meetings. Initial effect of attending projects' selection meetings on income growth is captured by the good performance of CDD at the commune level. We then explore the potential channels that participation might affect participants' final outcomes through different intermediate outcomes. Households, who participate, benefit from improved access to sanitary toilet, mobile phone, agricultural information and credit access. In particular, ethnic minorities and poor households can obtain higher gain from participation regarding these intermediate outcomes. At the commune level, CDD is found to be positively associated with higher economic outcomes and high school net enrolment.

Our findings highlight the benefit of active participation in infrastructure projects and one of the possible channels for this effect is through getting better information access. Improvement of rural roads has significant impacts on the development of local markets in Vietnam (Mu and van de Walle, 2011), thus, active participation in infrastructure project (mainly in road construction project) could create more opportunities for the participants. Nevertheless, we cannot rule out the bias in assessing participation approach in Vietnam given the local context. Is it actually the "quotas" imposed by local officials to meet their targets? Did participation approach actually happen in the field? Even if we cannot give a definite answer to those questions, our paper, to some extent provide suggestive evidence of positive effect of participation, at both household and commune level. Promoting active participation is needed, especially for the poor and ethnic minority since our finding shows that, having initially lower endowments, they can benefit more from participation. This study is limited in evaluating the impact of participation since we do not have data on specific project's outcomes. Future research could explore more the impact of participation at project level in Vietnam as well as investigate on political and institutional issues at local level to understand under which condition CDD and participation approach works best.

**References**


Ananga, E. O. 2015. 'The Role of Community Participation in Water Production and Management: Lessons From Sustainable Aid in Africa International Sponsored Water Schemes in Kisumu, Kenya.' Graduate Theses and Dissertations.

Andreoni, J. 1990. 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving?' *Economic Journal,* 100 (401): 464-77.

Arcand, J.-L., and Bassole, L. 2007. 'Does Community Driven Development Work? Evidence from Senegal'. Working Paper CERDI.

Bandiera, O., and Rasul, I. 2006. 'Social Networks and Technology Adoption in Northern Mozambique.' *The Economic Journal,* 116(514): 869-902.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. 2010. 'Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India'. *American Economic Journal*, Economic Policy, 2(1): 1-30.

Beard, V., and Dasgupta, A. 2006. 'Collective Action and Community Driven Development in Rural and Urban Indonesia.' *Urban Studies,* 43(9): 229–249.

Chebil, S., and Haque, I. 2003. 'Community Driven Development Programs for Poverty Reduction: Experiences, Issues, and Lessons'. *Scientific Journal of Administrative Development,* 1(3): 113-41.

Cling, J. P., Razafindrakoto, M., and Roubaud, F. 2003. 'The New International Poverty Reduction'. London/New York: Routledge.

Coleman, J., 1988. 'Social Capital in the Creation of Human Capital.' *The American Journal of Sociology,* 94: 95-120.

Culas, C., Razafindrakoto, M., & Roubaud, F. 'Ownership and Participatory processes: from global motto to local challenges. The case of a key poverty reduction program in ethnic minority areas of Vietnam.' In *Local Politics, Global Impacts: Steps to a Multi-Disciplinary Analysis of Scales*. O. Charnoz, V. Diaz Pedregal, A. L. Kolata (eds.). Surrey, UK: Ashgate Press: 223-250.

Cuong, N. V., Tung, P. D., and Westbrook, D. 2015. 'Do the poorest ethnic minorities benefit from a large-scale poverty reduction program? Evidence from Vietnam.' *Quarterly Review of Economics and Finance* 56: 3–14.

Do, N. 2008. 'Decentralization Process In Vietnam. Assessing Commune Investment Ownership under the Program 135'. A study of two communes, Kim Boi District, Hoa Binh Province, with UNDP/CEM Project Vie/02/001. Hanoi.

Dongier, P., Domelen, J., Van, Ostrom, E., Ryan, A., Wakeman, W., Bebbington, A., Alkire, S., Polski, M. (n.d.). Chapter 9 in PRSP sourcebook. Community-Driven Development: 301–31. Washington DC: The World Bank

Fafchamps, M., and Gubert F. 2007. 'The Formation of Risk Sharing Networks. ' *Journal of Development Economics* 83 (2): 326–50.

Gillespie, S. 2004. 'Scaling up Community-Driven Development: A synthesis of experience.' Washington, DC: World Bank.

Herrera J., Le Dang T., Pham H., Razafindrakoto M., Roubaud F. 2008. 'Impact evaluation of the Program P135 Phase 2: Analysis of the baseline survey 2007.' Hanoi: Committee for Ethnic Minorities Affairs (CEMA) and UNDP.

Karlan, D. 2007. 'Social connections and group banking.' *The Economic Journal,* 117 (37): F52-F84.

Khwaja, A. I. 2004. 'Is Increasing Community Participation Always a Good Thing?' *Journal of the European Economic Association*, 2(2)-3: 427-36.

Labonne, J., and Chase, R. S. 2009. 'Who is at the Wheel When Communities Drive Development? Evidence from the Philippines.' *World Development*, 37(1), 219–231.

Mansuri, G.; Rao, V. 2013. 'Can Participation Be Induced? Some Evidence from Developing Countries.' *Critical Review of International Social and Political Philosophy,* 16(2).

Mosse, D. 2001. 'People's Knowledge,' Participation and Patronage: Operations and Representations in Rural Development.' In Bill Cooke and Uma Kothari, eds., Participation: The New Tyranny. London: Zed Books.

Mu, R. and D. Van de Walle. 2011. 'Rural roads and local market development in Vietnam.' *The Journal of Development Studies* 47(5): 709–734.

Munshi, K. 2004. 'Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution'. *Journal of Development Economics* 73(1): 185-213.

Narayan, D. 1995. 'The Contribution of People's Participation: Evidence from 121 Rural Water Supply Projects'. ESD Occasional Paper Series, World Bank, 1(1).

Neef, A., Ekasingh, B., Friederichsen, R., Becu, N., Lippe, M., Sangkapitux, C., Gypmantasiri, P. 2013. 'Participatory Approaches to Research and Development in the Southeast Asian Uplands: Potential and Challenges'. In Sustainable Land Use and Rural Development in Southeast Asia: Innovations and Policies for Mountainous Areas.

Nkonya, E., Phillip, D., Mogues, T., Pender, J., and Kato, E. 2012. 'Impacts of Community-driven Development Programs on Income and Asset Acquisition in Africa: The Case of Nigeria'. *World Development*, 40(9): 1824–1838.

Nguyen, C., Phung, T., Phung, T., Vu, N., and Westbrook, D. 2013. 'The impact of a national poverty reduction program on ethnic minorities in Vietnam: The lens of baseline and end line surveys.' MPRA Paper No. 50477.

Nguyen, H.-T.-M., T. Kompas, T. Breusch, and M. B. Ward. 2017. 'Language, Mixed Communes, and Infrastructure: Sources of Inequality and Ethnic Minorities in Vietnam.' *World Development*, 96: 145-162.

Phung, D. T., Nguyen V. C., Phung, T., Vu, N., Le, D. T., Pham, T. H., Nguyen, T. N., Westbrook, D., and Taylor, J. 2012. 'Impact of Program 135 - phase II through the Lens of Baseline and Endline surveys'. Hanoi: UNDP.

Platteau, J-P., and Gaspart, F. 2003. 'The Risk of Resource Misappropriation in Community-Based Development Projects. ' *World Development*, 31(10): 1687-1703.

Prokopy, L. 2005. 'The relationship between participation and project outcomes: Evidence from rural water supply projects in India'. *World Development*, 33(11): 1801–1819.

Shanks, E., Toai, B. D., Nguyet, N. T. K., Maxwell, O., and Hung, D. Q. 2003. 'Community driven development in Vietnam'. Hanoi & Washington, DC: The World Bank & The Partnership to Support the Poorest Communes, Ministry of Planning and Investment.

UNDP. 2006. 'Vietnam policy dialogue paper. Deepening democracy and increasing popular participation in Vietnam.' Hanoi: UNDP.

Wassenich, P., and Whiteside, K. 2004. 'CDD impact assessment study: optimizing evaluation design under constraints'. Washington DC: Social Development Family, Work Bank.

White, H., Menon, R., and Waddington, H. 2018. 'Community-driven development: does it build social cohesion or infrastructure? A mixed-method evidence synthesis'. Technical Report. International Initiative for Impact Evaluation (3ie), New Delhi.

World Bank. 2009. 'Country Social Analysis : Ethnicity and Development in Vietnam - Main Report'. Hanoi: World Bank.

## Appendix

## Appendix 1

Definition and calculation method of key response variables

*Household asset index:* $\sum_{i=1}^{K} a_i$ where $a_i = 1$ if household has at least 1 asset i, and 0 otherwise

List of assets is in section 5 of the household questionnaires. They are production fixed assets such as rice milling machines, breeding castle, crop garden, etc.

*Household durable index*: $\sum_{j=1}^{K} b_j$ where $b_j = 1$ if household has at least 1 durable i, and 0 otherwise.

List of durables is in section 5 of the household questionnaires. They are household durables such as cookers, telephones, bikes, furniture, etc.

## Appendix 2

**Table A2 Intermediate outcomes and self-assessment by three measures of household participation**

*Attending project selection meetings*

| | 2007 | | | 2012 | | |
|---|---|---|---|---|---|---|
| Observations | Participants | Non-Participants | p-value | Participants | Non-Participants | p-value |
| Electricity | 0.763 | 0.652 | 0.00 | 0.856 | 0.774 | 0.00 |
| Piped or well water for drinking | 0.478 | 0.427 | 0.00 | 0.439 | 0.425 | 0.26 |
| Sanitary toilet | 0.159 | 0.136 | 0.05 | 0.386 | 0.294 | 0.00 |
| Mobile phone | 0.069 | 0.056 | 0.15 | 0.732 | 0.658 | 0.00 |
| Motorbike | 0.525 | 0.419 | 0.00 | 0.756 | 0.666 | 0.00 |
| Bike | 0.440 | 0.314 | 0.00 | 0.277 | 0.225 | 0.00 |
| Access to agricultural information | 0.434 | 0.398 | 0.02 | 0.655 | 0.477 | 0.00 |
| Access to credit | 0.596 | 0.526 | 0.00 | 0.695 | 0.584 | 0.00 |
| *Self-assessment* | | | | | | |
| Cash shortage | 0.682 | 0.735 | 0.00 | 0.796 | 0.838 | 0.00 |
| Water access shortage | 0.489 | 0.562 | 0.00 | 0.393 | 0.499 | 0.00 |
| Shortage in education payment | 0.412 | 0.340 | 0.00 | 0.520 | 0.515 | 0.77 |
| Shortage in medical payment | 0.284 | 0.298 | 0.32 | 0.314 | 0.327 | 0.35 |

*Expressing opinions in selection meetings*

| | 2007 | | | 2012 | | |
|---|---|---|---|---|---|---|
| Observations | Participants | Non-Participants | p-value | Participants | Non-Participants | p-value |
| Electricity | 0.798 | 0.687 | 0.00 | 0.882 | 0.798 | 0.00 |
| Piped or well water for drinking | 0.518 | 0.436 | 0.00 | 0.477 | 0.420 | 0.26 |
| Sanitary toilet | 0.185 | 0.138 | 0.00 | 0.451 | 0.313 | 0.00 |
| Mobile phone | 0.089 | 0.056 | 0.00 | 0.773 | 0.676 | 0.00 |
| Motorbike | 0.593 | 0.443 | 0.00 | 0.798 | 0.690 | 0.00 |
| Bike | 0.484 | 0.354 | 0.00 | 0.278 | 0.245 | 0.04 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Access to agricultural information | 0.477 | 0.400 | 0.02 | 0.701 | 0.535 | 0.00 |
| Access to credit | 0.623 | 0.547 | 0.00 | 0.718 | 0.621 | 0.00 |
| *Self-assessment* | | | | | | |
| Cash shortage | 0.647 | 0.723 | 0.00 | 0.763 | 0.830 | 0.00 |
| Water access shortage | 0.455 | 0.541 | 0.00 | 0.393 | 0.458 | 0.00 |
| Shortage in education payment | 0.414 | 0.369 | 0.00 | 0.538 | 0.512 | 0.16 |
| Shortage in medical payment | 0.267 | 0.297 | 0.08 | 0.312 | 0.323 | 0.53 |

*Contributing to infrastructure projects*

| | 2007 | | | 2012 | | |
|---|---|---|---|---|---|---|
| Observations | Participants | Non-Participants | p-value | Participants | Non-Participants | p-value |
| Electricity | 0.728 | 0.705 | 0.14 | 0.817 | 0.815 | 0.90 |
| Piped or well water for drinking | 0.422 | 0.469 | 0.01 | 0.390 | 0.450 | 0.00 |
| Sanitary toilet | 0.134 | 0.155 | 0.10 | 0.359 | 0.334 | 0.11 |
| Mobile phone | 0.052 | 0.068 | 0.07 | 0.714 | 0.688 | 0.08 |
| Motorbike | 0.505 | 0.464 | 0.03 | 0.751 | 0.696 | 0.00 |
| Bike | 0.426 | 0.363 | 0.00 | 0.270 | 0.244 | 0.06 |
| Access to agricultural information | 0.450 | 0.402 | 0.01 | 0.672 | 0.523 | 0.00 |
| Access to credit | 0.582 | 0.555 | 0.12 | 0.698 | 0.616 | 0.00 |
| *Self-assessment* | | | | | | |
| Cash shortage | 0.719 | 0.701 | 0.27 | 0.803 | 0.822 | 0.12 |
| Water access shortage | 0.497 | 0.534 | 0.03 | 0.406 | 0.462 | 0.00 |
| Shortage in education payment | 0.447 | 0.347 | 0.00 | 0.543 | 0.506 | 0.02 |
| Shortage in medical payment | 0.309 | 0.282 | 0.09 | 0.313 | 0.324 | 0.43 |

## Appendix 3

**Table A3 MCA statistics of columns – CDD score**

| | | | overall | | Dimension_1 | | |
|---|---|---|---|---|---|---|---|
| | Categories | mass | quality | %inert | coord | sqcorr | contrib |
| Meeting is organized | No | 0.047 | 0.890 | 0.197 | 2.075 | 0.889 | 0.201 |
| | Yes | 0.153 | 0.890 | 0.060 | -0.634 | 0.889 | 0.062 |
| Public bidding | No | 0.118 | 0.872 | 0.115 | 0.988 | 0.871 | 0.115 |
| | Yes | 0.082 | 0.872 | 0.164 | -1.410 | 0.871 | 0.164 |
| Community supervision board | No | 0.064 | 0.841 | 0.214 | 1.803 | 0.840 | 0.207 |
| | Yes | 0.136 | 0.841 | 0.100 | -0.840 | 0.840 | 0.096 |
| Household contribution | No | 0.179 | 0.951 | 0.009 | 0.228 | 0.919 | 0.009 |
| | Yes | 0.021 | 0.951 | 0.075 | -1.939 | 0.919 | 0.079 |
| Project financial information | No | 0.191 | 0.914 | 0.003 | 0.123 | 0.866 | 0.003 |
| | Yes | 0.009 | 0.914 | 0.065 | -2.739 | 0.866 | 0.064 |
| Types of project | | | | | | | |
| Commune's traffic road | 1 | 0.141 | 0.004 | 1.850 | 0.136 | 0.001 | |
| Village's traffic road | 2 | 0.157 | 0.056 | 1.821 | -0.788 | 0.047 | |
| Irrigation | 3 | 0.116 | 0.005 | 1.852 | -0.271 | 0.004 | |
| Market | 4 | 0.031 | 0.020 | 1.840 | 1.156 | 0.020 | |
| School | 5 | 0.155 | 0.002 | 1.823 | 0.034 | 0.000 | |
| Village art center | 6 | 0.087 | 0.051 | 1.755 | -0.917 | 0.036 | |
| Medical center | 7 | 0.079 | 0.013 | 1.832 | 0.551 | 0.011 | |
| Clean water | 8 | 0.102 | 0.004 | 1.814 | 0.280 | 0.004 | |
| Electricity line | 9 | 0.062 | 0.017 | 1.862 | 0.753 | 0.016 | |
| Local Committee office | 10 | 0.070 | 0.021 | 1.823 | 0.790 | 0.021 | |

# Chapter 2

# Heterogeneity in earnings gaps by gender and ethnicity:

## Evidence from Vietnam

Co-authored with François Roubaud

**Abstract**

We assess earnings gaps in Vietnam by gender and ethnicity, using the Labour Force Survey. The earnings gap between ethnic groups is larger than the earnings gap by gender and results are heterogeneous amongst different ethnic minority groups. We further employ a multinomial logit model to correct for selection and find that the gaps depend significantly on the sector choice. Selection is thus playing a key role in the Vietnamese labour market. Looking at the distribution of the earnings gaps, we find that the gender earnings gap is smallest at the bottom deciles of earnings while the ethnic earnings gap is largest at the bottom deciles of the distribution and is reducing with an increase in earnings. Using decomposition technique, we show that differences in sectoral allocation play an important role in explaining earnings gaps by ethnicity while gender gaps are observed mainly within sector. Lastly, unexplained earnings gaps still remain for both ethnic and gender gaps after being corrected for selectivity.

*Key words: Vietnam, employment, gender earnings gaps, ethnic earnings gaps, discrimination*

*JEL classification: J31, J71, O53*

**Acknowledgements:** The current version of this paper has benefited from valuable comments at the 6th GREThA International Conference on Economic Development, Bordeaux, France; the 9th Vietnam Economist Annual Meeting, Danang, Vietnam, and the DIAL-CEPED Workshop on "Inequality and Vulnerability in Vietnam" in Paris, France.

# 1 Introduction

Over the last 20 years, Vietnam has made outstanding progress in fighting poverty; however, economic growth has not affected all groups equally and Vietnam is now facing problems of "inclusiveness", i.e. making sure everybody contributes to and benefits from economic growth. One of the key obstacles to achieving inclusive growth in Vietnam is horizontal inequality or inequality between groups, in particular between men and women and between ethnic minority and majority groups. While Vietnam has achieved a notable reduction in gender inequality in aspects such as education, health, and female labour force participation, women are still more likely to be in vulnerable employment, as evident from their higher participation in the informal sector (World Bank, 2011). Moreover, women are over-represented in State owned enterprises and are therefore more strongly affected by the downsizing process of these enterprises. In terms of ethnic differences, although Vietnam is less diverse than other countries in the region, the country is characterized by a significant level of ethnic diversity. According to the last population Census (2009), 86 percent of Vietnam's population is Kinh (the majority group), while the rest of the population belongs to one of the 53 remaining ethnic groups officially registered in the country. Historically, ethnic minorities are mainly concentrated in mountainous and remote areas. While at the aggregate level, Vietnam records one of the best performances in the world in terms of economic growth and poverty reduction during the last two decades, ethnic minorities seem to have benefitted less than others to this ongoing process. Ethnic minority groups account for an increasing share of the poor in Vietnam. In 2010, the poverty rate was 66 percent among ethnic minority, compared to around 13 percent among the Kinh majority.

This diverging trend questions the inclusiveness of the growth pattern and its long term viability in terms of social cohesion. Persistent horizontal inequality has the potential to trap individuals and whole groups in poverty. Potential causes of the divergent growth pattern could be differences in economic characteristics between men and women or ethnic minority and majority groups; however, given equal productive endowments and location, discrimination could be another explanation. However, there is no absolute distinction between endowments and returns to endowments which is widely attributed to discrimination. The literature has shown that past discrimination could induce differences in current endowments between groups (van de Walle and Gunewardena, 2001). The context of past ethnic discrimination in Vietnam provides an interesting case to examine the issue. In this study, we look at the earnings gaps by gender and ethnicity in Vietnam and examine whether earnings discrimination exists in the labour market. A deeper understanding of the divergent growth pattern is crucial for the design of appropriate public policies by the Vietnamese government.

Numerous studies have been undertaken in the past years by independent academics or commissioned by the authorities, however, prior research only focuses on wage gaps and aggregated ethnic minority classification. This lack of research arises from the absence of suitable data on ethnicity and earning categories. Thus, our paper will contribute to the literature by addressing the earnings gaps by gender and ethnicity, taking advantage of a relatively comprehensive data in terms of ethnicity and earning categories, the Labour Force Survey (LFS), conducted by the General Statistics Office of Vietnam (GSO) in 2007 with technical assistance by one of the authors.

Our paper is organized as follows: section 2 places our paper within the existing literature, section 3 describes the data and provides an overview and main stylized facts concerning gender and ethnic groups' participation in the Vietnamese labour market. Then, section 4 describes our methodology to estimate different earnings gaps using strategies to tackle self-selection and

endogeneity biases. The results from our econometric estimations and robustness checks are discussed in section 5 and 6, respectively; section 7 concludes and draws some perspectives for further research.

## 2 Literature review

A long-standing literature on the economics of discrimination has indicated that two major theoretical models of discrimination in the labour market: taste-based discrimination and statistical discrimination. The former is initiated by the work of Becker (1971), whose model sees discrimination as a personal prejudice, or (dis-)taste, against a particular group or individual. There is no productivity-related attribute in the employment process: workers from a less favoured group simply have to accept lower wages or perform at higher productivity compared to members of the favoured group as a result of employers' preferences. In the market, taste-based discrimination can be seen when a customer is willing to buy a good at a relatively higher price at a preferred seller compared to purchasing it at cost from a less-preferred seller. The second theoretical model of discrimination is proposed by Phelps (1972) and Arrow (1973) who focus on statistical discrimination. In their models, due to asymmetric information, employers cannot identify workers' quality. Firms or employers then use easily observable characteristics such as ethnicity or gender to predict the expected productivity of applicants. In other words, based on stereotypes, employers judge a prospective employee by their own imperfectly measured ability and performance, as well as by that of their group (group average).

Empirically, a rich literature exists demonstrating that unequal treatment of women in the labour market it is not a country-specific phenomenon. Nopo et al. (2011) use data for sixty-four countries, ranging from developed to developing countries (including Vietnam) to provide a global picture of the gender earnings gaps. After controlling for observable socio-demographic and job characteristics, Nopo et al.'s estimate for the gender earnings gaps range from 8 percent to 48 percent of average women's earnings. Typically, the unexplained part of the earning gaps is larger for people having only part time jobs and low education levels. Literature further focuses on the determinants of the gender earnings gaps, such as differences in wage structure and women's labour supply (Blau and Kahn, 2003), sectoral allocation (Tzannatos, 1999; Nordman et al., 2011), liberalization policies in the labor market and country-specific institutional frameworks (Weichselbaumer et al., 2008; Blau and Kahn, 2003 and Tzannatos, 1999). The unexplained parts of the earnings gaps could be due to unobservable factors and/or discrimination against women.

Studies on the gender earnings gaps in Vietnam have been largely focused on changes in women's participation in the labor force and inequality in wages between male and female workers. Gallup (2004) uses two rounds of the Vietnam Household Living Standards Survey (VHLSS) in 1992-93 and 1997-98 and finds that women in Vietnam earn less than men with the same observable characteristics, although the differences in wages decrease between 1993 and 1998. Using the same data, Liu (2004a, 2004b) provides sectoral decompositions and concludes that the gender pay gap is highest in the private sector, followed State-owned enterprises (SOEs) and lowest in the public sector. By decomposing the gap, she finds that although the gender pay gap has narrowed over time, discrimination against women in the labour market remains high and is driven by within rather than between sector differences. Pham and Reilly (2007) explore three VHLSS rounds (1993, 1998 and 2002) to investigate the gender wage gap through mean and quantile regression methods. They observe that the gender wage gap reduces by half over the 1993-2002 period and most of the achievement is observed in the 1993 – 1998 period. The

gap is found narrowing at the bottom end and most sizeable at the top end of the conditional wage distribution.

Fewer studies have focused on exploring ethnic earnings gaps, likely due to data limitations, and results on ethnic earnings gaps across the world have so far been inconclusive. On the one hand, Atal et al. (2009) highlight that in Latin American countries, the ethnic wage gap is larger than the gender wage gap and differences in human capital and occupational segregation are key explanations for the ethnic wage gaps in the region. Examining the structure of wage gaps by ethnicity and gender in Bolivia, Ecuado and Guatemala, Canelas and Salazar (2014) also show that women are highly discriminated while ethnic gaps are also large but mainly explained by the endowment differences. On the other hand, studies on some African countries reveal low and mainly insignificant ethnic earning gaps (Roubaud, 1994; Nordman et al., 2011). In the case of Vietnam, Pham and Reilly (2009) use the VHLSS 2002 and find that the ethnic wage gap, is substantially smaller than the gap observed by other researchers using household living standard measures (e.g. Van de Walle and Gunewardena, 2001; Baulch et al., 2007). Using quantile regression models, Pham and Reilly (2009) reveal a gap of nearly 21 percentage points at the bottom decile compared to only four percentage points at the top deciles. The authors attribute the ethnic wage gap is mainly to the differentials in returns to endowments. However, the authors do not examine the ethnic wage gap for the self-employed and the gap is considered at aggregated level between Kinh and Non-Kinh, failing to take into account the heterogeneous composition of the Non-Kinh group. Baulch and Vu (2011) focus on expenditures instead of earnings. Using the VHLSS 2004 and 2010 to analyze the factors driving the ethnic expenditure gap, they show an increase in the ethnic expenditure gap from 58.1 percent in 2004 to 68.1 percent in 2010. Differences in endowments only explain a small part of the gap (two fifth of the gap in 2004 and nearly half in 2010). Again, differences in returns are the more important component, accounting for more than half of the expenditure gap, implying that ethnic minorities are likely to be discriminated.

While considerable evidence exists on the earnings gaps by gender and ethnicity in Vietnam, these papers exhibit important shortcomings to necessitate the analysis conducted in this paper. First, these papers typically only focus on wage employment or household expenditures and thereby fail to account for the self-employment and those working in the agricultural sector. However, wage employment only account for a small fraction of the labour force in Vietnam (approximately 30 percent in 2007, 2009) and previous results should therefore be interpreted with caution.

Second, previous papers have not fully accounted for selection into wage work. Liu (2004a, b) focuses on the differences in returns of a wage-worker in one sector from another. She uses the traditional Oaxaca and Neumark decomposition and additionally Appleton et al. (1999)'s method to correct for the differences attributed to sectoral choice differentials. However, she does not correct for the two types of sample selection bias due to the fact that 1) not everyone is working and 2) thee selection into wage employment. Liu (2004a, b) then only uses a multinomial logit model (Lee, 1983) in the first step to correct for the selectivity between three different types of wage employment, which is insufficient. Pham and Reilly (2007, 2009) employ the same method.

Third, no previous paper has looked at subgroups of the ethnic minorities, instead considering

them one homogenous group. It is possible though, that people from one ethnic minority group could earn equal or even higher incomes than the majority, which we could not realize when the whole 53 ethnic minority groups are aggregated.

In Vietnam, ethnic minorities are mostly located in rural and remote areas. Given their remote location, ethnic minorities are mostly employed in subsistence farming or in self-employed activities. Focusing on the wage gap without properly accounting for selection into wage employment would thus heavily underestimate the real gap between the ethnic majority and minority. Similarly, as women tend to be overrepresented in public sector (Liu, 2004b) which traditionally pays lower wages, focusing only on wage gaps would underestimate the gender wage gap.

This study examines the gender and ethnic earnings gaps across different employment sectors, by first, taking into account self-employed workers and farmers; second, properly correcting for selection bias; and third, providing disaggregated estimates for the 53 different ethnic minority groups. We will start with analyzing different earnings gaps: the gender and ethnic earnings gaps, non-farm earnings gaps, and wage gaps, and then decomposing these gaps in different employment sectors while correcting for selection bias. We will look at ethnic groups both as a minority aggregation and disaggregating the group into more detailed ethnic minority groups.

## 3 Data and descriptive statistics

### 3.1 The data and ethnic group's aggregation

Our study is based on the Vietnam Labour Force Survey 2007 (LFS2007) conducted by the GSO in August 2007. The LFS2007 remains the best source of data to address the ethnic earnings gap in Vietnam and more broadly issues related to ethnic groups in the labour market. Previously conducted every year by the Ministry of Labour Invalids and Social Affairs (MoLISA), the Vietnam LFS has been transferred to the GSO in 2007. Taking advantage of this institutional transfer, the survey has been integrally revamped (sample design and questionnaire) with the technical support of the authors. The number of questions has been increased from around 25 to 50, allowing for an enlarged coverage of labour market indicators and data quality has been substantially improved (Razafindrakoto and Roubaud, 2007; GSO, 2008).

Three key points of the LFS upgrading makes this survey perfectly suited for this study. First, the scope of survey has been enlarged. Initially designed to capture only official unemployment, the LFS2007 embraces many other facets of the labour market. In particular, it captures the informal sector. In Vietnam, the informal sector is defined as all private unincorporated enterprises that produce at least some of their goods and services for sale or barter, are not registered (do not have a statistics license which is compulsory for all kinds of businesses) or do not keep book accounts. In this paper, we distinguish between four labour market segments: public sector, formal private sector, informal sector, and agriculture. It is noted that this sector classification is totally based on institutional sector, without taking into account employment status, such as wage worker or self-employed. Apart from the formal/informal sector divide, the survey dedicated special care to obtain reliable measures of variables for which the informality status may lead to sampling and measurement errors. In particular, the questionnaire includes a detailed set of questions to capture information on activity status, the classical procedures leading to the under-declaration of informal sector workers' participation for those with the weakest labour market attachment.

Second, the LFS2007 captures a wide range of labour earnings. We compute the labour income associated with each remunerated job (in main and secondary occupation). For wage workers, the survey captures their current monthly wage, while for self-employed workers earnings correspond to the disposable income (before taxation). Measuring non-wage workers' earnings is all complicated by the fact that employees only represent a small fraction of the labor force (see below). The LFS2007 is the only survey in Vietnam to gather information on self-employed earnings, an issue which has not been considered in later rounds of Vietnamese LFS. Hourly earnings used in the econometric analysis are constructed using the total number of hours worked per month. In order to account for differences in prices levels by location, we use the deflators drawn from the VHLSS2006, to compute earnings in constant 2006 Vietnam Dongs. 16 sets of prices are distinguished, for each domain of inference of the VHLSS2006: the 8 official regions by area (urban/rural). Using the LFS2007 thus allows us to investigate the earnings gaps in the informal sector and self-employed whereas previous analyses had to be confined to wage workers only.

Third, the LFS2007 is a unique dataset to investigate labour market issues related to ethnicity in two aspects: individual ethnic affiliation and sample size. Vietnam's official classification of ethnic minorities considers 54 different ethnic groups and LFS2007 is the only survey to allow individuals to sort themselves into any of the 54 different groups. . More recent rounds of the LFS do not contain disaggregated information on ethnic groups but instead only allow a disaggregation into the six major ethnic minority groups (Tày, Thái, Muong, Kho-me, Nung, Hmông) with remaining individuals being classified as 'others'.

According to the LFS2007 (Table 1), 86 percent of the population is ethnically Kinh (the majority group) and the remaining 14 percent of the population belongs to ethnic minorities, representing 11.2 million individuals. Among the 53 minority ethnic groups, the distribution is highly skewed. Some groups account for more than one million individuals (Thái, Tày, Muong, Kho-me), others account for almost one million (Nùng, Hoa), smaller ones represent only a few thousands (Pu Peo, Bo Y, Co Lao, etc.) or even a few hundred people (Brâu, Si La). Of course, as in other countries, the official ethnic classification is partly subjective. Historical and anthropological works have shown the determination of ethnic groups is a construct and political process in Vietnam (McElwee, 2004). Ethnicity is a social construct and ethnic identity is embedded in a political process elaborated in the long run. Notwithstanding these reservations, the paper will use the classification obtained in the LFS2007 for the following reasons. First, this classification is widely accepted by the population itself: based on self-declaration, non-response concerns less than 15 observations. Second, many economic and social policies have been designed according to this ethnic classification in Vietnam, producing effects that could only be assessed when sticking to the same classification.

**Table 1: Sample size by ethnic groups in the LFS2007**

| | Population | | Structure | |
|---|---|---|---|---|
| | **Sample size** | **Number (1,000)** | **Total** | **Without Kinh** |
| **Kinh** | 561,570 | **72,585** | **85.9%** | **-** |
| **Non Kinh** | 99,710 | **11,951** | **14.1%** | **-** |
| Hoa | 7,338 | 939 | 1.1% | 7.9% |
| Thái | 13,611 | 1,989 | 2.4% | 16.6% |
| Tày | 22,429 | 1,745 | 2.1% | 14.6% |
| Muong | 6,722 | 1,325 | 1.6% | 11.1% |
| Kho-me | 6,944 | 1,095 | 1.3% | 9.2% |
| Nùng | 9,276 | 946 | 1.1% | 7.9% |
| Hmông | 8,490 | 831 | 1.0% | 7.0% |
| Dao | 4,986 | 510 | 0.6% | 4.3% |
| Ê Đê | 2,066 | 363 | 0.4% | 3.0% |
| Gia-rai | 2,282 | 326 | 0.4% | 2.7% |
| Ba na | 1,909 | 181 | 0.2% | 1.5% |
| Xo-dang | 2,228 | 184 | 0.2% | 1.5% |
| Ra-glai | 1,551 | 170 | 0.2% | 1.4% |
| Co-ho | 977 | 159 | 0.2% | 1.3% |
| Chăm | 1,126 | 145 | 0.2% | 1.2% |
| Giáy | 1,323 | 124 | 0.1% | 1.0% |
| Other Northern minorities | 2,342 | 365 | 0.4% | 3.1% |
| Other Central Highland | 3,465 | 412 | 0.5% | 3.4% |
| Others | 645 | 142 | 0.2% | 1.2% |
| **Total** | **666,280** | **84,536** | **100%** | **100%** |

Source: LFS2007, GSO; authors' calculations.

Ethnic identification in the LFS2007 is recorded at the individual level. Each member of the household is supposed to declare her own ethnic affiliation, whereas previous surveys only collected data for the head of household, relying on the implicit hypothesis that all households are endogamous. This seems to be an erroneous assumption as Census data shows high levels of intermarriage for some ethnic groups. Even though within the total population only 2 percent of couples belong to different ethnic groups, among ethnic minorities, 15 percent are exogamous couples (1 percent for the Kinh). The propensity to exogamy varies greatly with the ethnic groups: from only a few percent for the Thái, the Muong, the Gia Rai, the Co-ho, the Chăm, etc., to much higher proportions among the Tày (22 percent), the Hmông (25 percent), the Kho-me (44 percent), and even 100 percent for the È Đê in our dataset. As the aim of this paper is to look at the individual earnings gap by ethnicity, reliable and detailed data on ethnicity is crucial.

The final reason for using the LFS2007 concerns its sample size. With 173,000 sampled households and 661,000 persons (100,000 individuals belonging to ethnic minorities, out of which 55,000 persons have a job), the LFS2007 provides a much higher potential level for ethnic groups disaggregation, properly reflecting ethnic heterogeneity in Vietnam. Most of the previous quantitative works on ethnic issues in Vietnam have focused on the separation between Kinh and Non-Kinh (Van de Walle and Gunewardena, 2001; Baulch et al., 2004). More recent papers have considered 8 ethnic groups: Kinh & Hoa, Kho-me & Chăm, Thái, Tày, Muong and Nùng, Other Northern minorities, Central Highland minorities, Others (Baulch et al., 2007, 2008, and 2010[6]). Due to the large sample size of the LFS2007, we can increase the number of ethnic categories while retaining statistical significance, which is crucial as differences between ethnic minority groups may be quite substantial, and in some cases, probably higher than differences observed between aggregate categories. In this paper, we group ethnic minorities according to sample size. With more than 1,000 observations, 17 ethnic groups form their own groups, while the remaining ethnic minorities with less than 1,000 observation in the sample, are grouped together into 3 groups aggregated by main location (*Other Northern minorities*, *Other Central Highland*, *Others*). In total, our typology distinguishes 20 different ethnic groups. More broadly and apart from ethnic disaggregation, thanks to its sample size, the LFS2007 provides more precise estimates for all indicators, at a higher level of inference (provincial vs. regional).

### 3.2 Gender and Ethnic minorities in the labor market: a descriptive analysis

The main labour market indicators computed from the LFS2007 offer evidences of strong differences between Kinh and ethnic minorities regarding and some difference between males and females (Table 2). First, the share of the working age population (equal or over 15 years old) is higher for women than men (76 percent vs. 73 percent), but the employment rate is higher for men by 8 percent. Women are more likely to have a second job than men, which reflects the fact that women might choose to have a temporary job in order to have time for taking care of their children and family. In terms of ethnicity, the working age population rate is significantly higher for Kinh than non-Kinh (76 percent vs. 69 percent), reflecting higher fertility and mortality rates (Friedman et al., 2001), a larger share of children, and a higher dependency ratio among ethnic minorities. Hoa are the only ethnic group to have a larger share of working age

[1] These papers use the VHLSSs, whose sample size is limited to around 9,000 households.

adults than Kinh.[7] Conversely, the activity rate is on average almost 10 percentage points higher for ethnic minorities (82 percent vs. 71 percent), due to lower school attendance (for the young generation) and a higher rate of households' labour force mobilization in primary sectors. Consistently, the rate of unemployment is lower for ethnic minorities (1 percent vs. 2 percent), providing evidence for the hypothesis of luxury unemployment on Vietnamese labour market stressed in other studies (Razafindrakoto et al., 2011), while the rate of multi-activity (workers holding a second job) is higher among non-Kinh than Kinh (23 percent vs. 17 percent), reflecting a higher attachment of ethnic minorities to farming activities. Surprisingly, the time related underemployment rate[8] in the main job does not differ between majority and minorities (around 11 percent for both), in spite of fewer working hours for the latter. This feature probably results from a lower propensity to declare wanting more jobs within ethnic minorities, because of scarce labor market opportunities. Another explanation is that ethnic minorities are likely to have more jobs, which means they might want to work more hours in the main job but cannot be able to.

**Table 2: Structure of the labor force by gender and ethnicity in 2007**

| | Working age population | Employment rate | Unemployment rate | Under-employment rate | Second job |
|---|---|---|---|---|---|
| **By gender** | | | | | |
| Males | 72.9 | 77.13 | 2.02 | 11 | 17.2 |
| Females | 76.1 | 69.31 | 1.9 | 10.7 | 19.2 |
| **By ethnicity** | | | | | |
| Kinh | 75.5 | 71.2 | 2.1 | 10.9 | 17.4 |
| Non-Kinh | 68.8 | 82.1 | 0.9 | 10.5 | 23.1 |
| Hoa | 79.8 | 62.0 | 4.4 | 5.6 | 3.1 |
| Thái | 69.4 | 86.7 | 0.4 | 11.6 | 25.0 |
| Tày | 74.5 | 80.3 | 1.0 | 9.8 | 32.3 |
| Muong | 73.2 | 84.2 | 0.5 | 10.7 | 14.5 |
| Kho-me | 72.2 | 76.2 | 1.7 | 16.9 | 16.7 |
| Nùng | 68.8 | 85.2 | 0.5 | 9.3 | 35.9 |
| Hmông | 55.0 | 92.3 | 0.1 | 10.7 | 35.8 |
| Dao | 64.6 | 88.0 | 0.3 | 3.0 | 28.5 |
| Ê Đê | 63.9 | 79.7 | 0.5 | 10.9 | 14.2 |
| Gia-rai | 59.1 | 90.8 | 0.4 | 10.5 | 21.6 |
| Ba na | 59.6 | 90.9 | 0.1 | 10.6 | 15.9 |
| Xo-dang | 56.0 | 91.2 | 0.2 | 26.8 | 37.3 |
| Ra-glai | 56.9 | 88.1 | 0.1 | 6.4 | 20.9 |
| Co-ho | 60.6 | 84.6 | 0.2 | 12.1 | 8.9 |
| Chăm | 64.6 | 81.5 | 1.4 | 9.0 | 7.9 |
| Giáy | 63.3 | 86.3 | 0.2 | 0.2 | 13.5 |
| Other Northern | 65.7 | 86.5 | 0.8 | 11.1 | 24.0 |
| Other Central H. | 64.0 | 79.1 | 0.8 | 12.3 | 22.0 |
| Others | 63.5 | 87.7 | 0.0 | 1.6 | 7.7 |
| **Total** | **74.6** | **73.0** | **2.0** | **10.8** | **18.2** |

Source: LFS2007, GSO; authors' calculations.

[2] The specificity of the Hoa group (ethnic Chinese), which we will encounter for almost all labor market indicators, leads to some authors preferring to consider Kinh and Hoa together and to differentiate between them and other ethnic groups. In this paper, we choose to identify Hoa separately, since Hoa are oftentimes better off than the rest of the population (including Kinh). As a result, if we were to consider Hoa and Kinh as one group, results for Kinh would be systematically overstated.

[8] Time related underemployment is the ratio of the occupied population working less than 35 hours per reference week and wanting to work additional hours.

For this set of indicators, the situation of Kinhs contrasts sharply with the ethnic minority groups (except for Hoa). We can perceive further differences among ethnic minority groups. Hmông have the lowest working age population share, 55 percent, but a rate of activity as high as 92 percent. The Hoa community, occupying the other end of the minorities' ladder, presents exactly the opposite, with the highest rates of working age population and unemployment (80 percent and 4 percent respectively), the lowest rates of employment (62 percent), and one of the lowest underemployment rate (6 percent).

**Table 3: Demographic characteristics of the labor force by gender and ethnicity in 2007**

| | Rural area (%) | Age | Year of schooling | Married (%) | Household size |
|---|---|---|---|---|---|
| **By gender** | | | | | |
| Males | 74.8 | 37.8 | 8.4 | 75.4 | 4.6 |
| Females | 75.4 | 38.3 | 7.8 | 73.6 | 4.5 |
| **By ethnicity** | | | | | |
| Kinh | 72.6 | 38.5 | 8.5 | 74.8 | 4.4 |
| Non-Kinh | 89.5 | 35.1 | 5.8 | 72.5 | 5.3 |
| Hoa | 21.9 | 38.5 | 7.5 | 57.0 | 4.4 |
| Thái | 97.0 | 34.3 | 5.8 | 75.8 | 5.3 |
| Tày | 91.3 | 35.7 | 8.1 | 74.3 | 5.5 |
| Muong | 97.3 | 35.6 | 8.0 | 72.8 | 4.6 |
| Kho-me | 88.5 | 36.5 | 4.1 | 68.0 | 4.9 |
| Nùng | 93.5 | 35.0 | 6.2 | 74.9 | 4.8 |
| Hmông | 98.6 | 32.9 | 2.0 | 77.5 | 5.0 |
| Dao | 97.3 | 33.0 | 3.8 | 79.0 | 6.6 |
| Ê Đê | 86.8 | 34.7 | 3.9 | 70.1 | 5.7 |
| Gia-rai | 94.1 | 34.7 | 3.6 | 69.6 | 6.1 |
| Ba na | 93.5 | 33.9 | 3.2 | 73.6 | 5.8 |
| Xo-dang | 96.6 | 34.9 | 4.1 | 69.9 | 5.5 |
| Ra-glai | 96.3 | 33.0 | 2.5 | 67.5 | 5.6 |
| Co-ho | 99.6 | 33.8 | 4.1 | 69.9 | 6.1 |
| Chăm | 89.5 | 35.0 | 4.5 | 72.5 | 6.4 |
| Giáy | 95.3 | 32.5 | 4.8 | 72.2 | 5.4 |
| Other Northern | 96.1 | 33.8 | 5.8 | 74.9 | 5.3 |
| Other Central H. | 93.4 | 34.6 | 4.1 | 70.7 | 5.6 |
| Others | 95.8 | 35.3 | 6.3 | 73.3 | 5.5 |
| **Total** | **75.1** | **38.0** | **8.1** | **74.5** | **4.5** |

Source: LFS2007, GSO; authors' calculations.

Looking at demographic characteristics of the labour force by gender and ethnic groups (Table 3), we can seen that women and men are living in urban and rural areas at similar rates and are on average the same age. However, women in the labour force have fewer years of schooling than men (7.8 vs. 7.4) and proportion of married women is also lower. Ethnic minorities are consistently more often located in rural areas than Kinh (90 percent vs. 73 percent), except for Hoa who are 78 percent urban. In fact, for most of the ethnic groups, the proportion of rural labour force participation exceeds 95 percent, except for Gia-rai (94 percent), Nùng and Other Central Highland minorities (93 percent), Tay (92 percent), Kho-me and Chăm (89 percent), and Ê Đê (87 percent). Ethnic minorities are younger (35 years old on average vs. 39 for the Kinh), and less educated than the majority. Kinh labour force has on average 8.5 years of schooling, while ethnic minorities have only 5.8 years. A fourth of Kinh labour force has completed higher secondary education compared to only a tenth for ethnic minorities, with 8

ethnic groups having less than 5 percent of their labour force with completed secondary schooling. Again, Hmông are in the worst situation (with 2 years of schooling on average and 1 percent of the labour force attended tertiary education). Tày, Muong and to a lesser extent, Hoa perform relatively well in this respect (8.1, 8.0 and 7.5 years of schooling each).

The structure of jobs by institutional sectors gives further insights to job allocation between men and women and among ethnic groups (Table 4). Men are more prevalent in the public and formal sectors while there is no difference between the share of men and women employed in the informal sector. Within the formal sector, women are engaged more than men in foreign enterprises but less in domestic enterprises and formal household business. Agriculture is the only sector where women are involved more than men. In terms of ethnicity, the massive participation of ethnic minorities in agricultural activities leads to a systematic underrepresentation in other institutional sectors. Given their underrepresentation in the private sector, ethnic minorities are even more underrepresented in the nonfarm private sector than in the public sector due to an active inclusion policy in the latter. Still, the share of ethnic minorities in public jobs is less than half the share for Kinh's (5 percent vs. 12 percent). Within the private sector, ethnic minorities are engaged in formal household business (3 percent vs. 9 percent for Kinh), followed by the informal sector[9] (7 percent vs. 26 percent) and domestic enterprises (2 percent vs. 6 percent). In this respect, foreign enterprises perform the worst: the relative chance for Kinh to get a job in foreign enterprises is more than five times higher than for ethnic minorities (2.2 percent vs. 0.4 percent), and twelve out of our nineteen ethnic groups have no representation in this sector. Foreign enterprises have been the fastest growing sector in the last decade and the sector which provides the best jobs in terms of labour conditions. The exclusion of ethnic minorities from foreign enterprises is thereby problematic for inclusive growth. It can probably be explained by the concentration of FDI to urban centers, (85 percent in Hanoi and Ho Chi Minh City surroundings; ADB, 2004; Cling et al., 2009), and limited skills availability among ethnic minorities (foreign enterprises tend to recruit mainly semi-skilled workers) rather than to explicit discrimination. Going into more details, evidence suggests that some ethnic groups seem to be keener to work in specific institutional sectors. This seems to be the case for Hoa who are more likely to be engaged in formal household businesses (30 percent of the labour force), compared to 8 percent of Kinh and 3 percent of other non-Kinh. Kho-me and Chăm seem to prefer the informal sector, while Central Highland minorities are globally excluded from the formal private sector. Tày (9 percent) and Kinh (12 percent) are best integrated into the public sector.

[9] According to GSO, the informal sector comprises all non-registered (no business license), non-farm household businesses. Formal household businesses refer to registered businesses.

**Table 4: Main job structure by institutional sectors, gender and ethnicity**

| | Public sector | Formal sector | | | Informal sector | Agriculture |
|---|---|---|---|---|---|---|
| | | Foreign enterprise | Domestic Enterprise | Formal HBs | | |
| **By gender** | | | | | | |
| Males | 11.2 | 1.5 | 6.9 | 8.2 | 23.8 | 48.4 |
| Females | 9.9 | 2.5 | 4.7 | 7.4 | 23.3 | 52.2 |
| **By ethnicity** | | | | | | |
| Kinh | 11.5 | 2.3 | 6.4 | 8.5 | 26.4 | 44.6 |
| Non-Kinh | 4.7 | 0.4 | 1.7 | 3.3 | 6.5 | 83.5 |
| Hoa | 6.8 | 3.4 | 12.3 | 30.1 | 31.2 | 15.8 |
| Thái | 4.11 | 0.0 | 0.8 | 0.7 | 2.4 | 92.0 |
| Tày | 8.9 | 0.1 | 1.4 | 2.1 | 4.8 | 82.7 |
| Muong | 5.0 | 0.2 | 0.5 | 1.3 | 4.7 | 88.1 |
| Kho-me | 3.5 | 0.6 | 2.7 | 4.0 | 18.3 | 70.9 |
| Nùng | 4.8 | 0.2 | 0.6 | 1.4 | 3.8 | 89.2 |
| Hmông | 0.9 | 0.0 | 0.1 | 0.1 | 0.4 | 98.5 |
| Dao | 2.2 | 0.0 | 0.1 | 0.3 | 2.1 | 95.2 |
| Ê Đê | 2.4 | 0.0 | 0.7 | 1.0 | 1.2 | 94.8 |
| Gia-rai | 2.9 | 0.0 | 0.0 | 0.0 | 0.9 | 96.3 |
| Ba na | 1.8 | 0.0 | 0.3 | 0.5 | 0.8 | 96.7 |
| Xo-dang | 1.9 | 0.1 | 0.0 | 0.4 | 0.0 | 97.3 |
| Ra-glai | 1.3 | 0.0 | 0.0 | 0.3 | 2.7 | 95.8 |
| Co-ho | 1.6 | 0.0 | 0.0 | 0.3 | 0.9 | 97.2 |
| Chăm | 3.9 | 0.1 | 2.7 | 0.3 | 9.3 | 83.8 |
| Giáy | 3.0 | 0.0 | 1.1 | 0.4 | 3.1 | 92.3 |
| Other Northern | 4.3 | 0.0 | 0.6 | 1.4 | 7.6 | 85.9 |
| Other Central H. | 5.7 | 0.0 | 0.2 | 0.6 | 1.6 | 91.9 |
| Others | 3.9 | 0.0 | 0.4 | 1.2 | 2.5 | 92.1 |
| **Total** | **10.5** | **2.0** | **5.7** | **7.7** | **23.5** | **50.3** |

Source: LFS2007, GSO; authors' calculations.

## 4 Empirical strategy

### 4.1 Earnings determinations

***Mincer earning equations***

In this paper, we estimate classical Mincerian earnings equations with different specifications to address potential self-selection and endogeneity biases and conduct several robustness checks of the results. We start from our simplest specification which is as follows:

$$\ln w_i = \beta x_i + \varepsilon_i \qquad (1)$$

Where $\ln w_i$ is the natural logarithm of the observed hourly real earnings for individual i, $x_i$ is a vector of observed characteristics for individual i, β is a vector of coefficients and $\varepsilon_i$ is a disturbance term with an expected value of zero. We calculate real hourly earnings, our dependent variable, to account for differences in consumer price levels by location using the deflators drawn from the VHLSS2006. The deflators distinguish between 16 different domains, namely the 8 official regions subdivided by area (urban/rural). We use a dummy for gender and dummies for ethnicity. Independent variables include productivity-related factors such as education and market experiences, with squared of those variables. Due to the data's nature, we do not have information about individual's actual experience but merely potential experience which is measured by the time that the individuals have spent in the labor force (calculated as individual's age minus years of education and minus six), without taking into account the interrupted experiences. Thus, the first regression is run on a basic Mincerian equation, including independent variables of marital status, years of schooling and potential experiences. Our second regression includes province fixed effects (64 provinces), to partially tackle endogeneity concerns and capture non-price local effects. In the third regression, we keep the province fixed effects and add independent variables of the institutional sector and the urban/rural divide. In terms of ethnic divide, models are estimated at the usual dichotomous level (majority vs. minorities or Kinh vs. Non-Kinh in our case) and then by using the full disaggregation of ethnic minorities (same specifications with 20 ethnic groups).

Our first regressions follow traditional studies on earning gaps in using a simple OLS regression for the earnings estimation, considering only actively occupied people. However, if a significant part of the population is not working, endogenous sample selection might be an issue if unobservable factors in the earnings equation and in the participation equation are correlated. An additional selection problem arises from the decision of workers to participate in different sectors, which are subject to different standards of pay. Indeed, besides the fact that earnings are only observed for people who work, people who choose to engage in the public sector have different characteristics than people who engage in the formal, informal, or agricultural sector, and those characteristics could affect their earnings gaps. Therefore, we address these two potential selection biases by using a multinomial selection model presented in Bourguignon et al. (2007, hereafter BFG), relying on a generalized version of the Dubin-McFadden correction [10].

[10] See Bourgignon et al. (2007) for a survey of the available methods to obtain consistent estimates of $\beta_s$ and $\gamma_s$ with two-step procedure.

***Multinomial logit model***

The multinomial selection model is given as:

$$y_s = X_s\beta_s + \epsilon_s \qquad (2)$$

$$y_s^* = Z_s\gamma_s + v_s \qquad (3)$$

Where s=1, 2, 3, 4, 5, reflecting the 5 alternatives from which each individual can choose: i) public sector employment; ii) formal sector employment; iii) informal sector employment; iv) agricultural sector employment and v) not working[11]. The vector Z represents the maximum set of explanatory variables for all alternatives and the vector X contains all determinants of earnings; $y_s$ is the earning corresponding to each employment choice presented by $y_s^*$ - a discrete-choice variable. As mentioned above, the estimation of $\beta_s$ in OLS regression will be inconsistent if the unobservable characteristics of individuals affect both their earning and choice of employment, thus if error terms $\epsilon_s$ and $v_s$ are correlated. Following the BFG methodology, we will estimate the earnings equation while including correction terms for sectoral choice. The correction model is based on the assumption of a linear association between $\epsilon_s$ and $v_s$ ($\epsilon_i = \sigma_i \sum_s \rho_s v_s$). The earnings equation corrected for potential bias is formed as follows:

$$y_1 = X_1\beta_1 + \sigma_1 \left[\rho_1 m(P_1) + \sum_s \rho_s \frac{P_s}{(P_s-1)} m(P_s)\right] + w_1 \qquad (4)$$

Where $P_s$ is the probability that a category $s$ is chosen, $w_1$ is a residual that is mean-independent of the regressors and $m(P_s) = \int J(w - logP_s)g(w)dw$. The number of bias correction terms is equal to the number of alternatives in the multinomial logit (thus 5 in our case).

In practice, we estimate the model in two stages. In the first stage, a multinomial logit model is estimated in which the predicted probabilities of each individual's participation in sector s are used to calculate the correction terms. To identify sectoral participation, we need an instrument which affects only the propensity to work and sectoral choices but is not correlated with the error term. The literature has suggested using household structure or demographic variables as exogenous determinants of sector choices. In line with the previous studies (Appleton et al., 1999; De Brauw and Rozelle, 2008; Pham and Reilly, 2007), we use relationship to the household head, household size and the dependency ratio (number of non-working age individuals divided by the total number of individuals in the household) as instruments. In the second stage, we then use these predicted correction terms to control for sectoral selection in the earnings equations. Having obtained credible estimates for the predictors of earnings, we now focus on the decomposition of the earnings gaps.

[11] Hausman-test was conducted to ensure the validity of Independence of Irrelevant Alternative (IIA) assumption

### 4.2 Earnings gaps decomposition

We begin with our decompositions with Oaxaca and Blinder decompositions. We estimate the earnings equations (equation 1) separately for Kinh and non-Kinh/ethnic minorities (aggregated together) and then use the Oaxaca-Blinder decomposition techniques to identify the sources of the ethnic earning gaps. In an analogous approach, we can decompose the gender earnings gaps by replacing Kinh with men and non-Kinh with women. The decomposition is then given by:

$$\overline{lnW_k} - \overline{lnW_n} = \beta_k(\bar{x}_k - \bar{x}_n) + (\beta_k - \beta_n)\bar{x}_n \quad (5)$$

where $w_k$ and $w_n$ are the average earnings for Kinh (or men) and ethnic minorities (or women), respectively; $\bar{x}_k$ and $\bar{x}_n$ are vectors of average observed characteristics for Kinh (or men) and minorities (or women); $\beta_k$ and $\beta_n$ are the estimated coefficients. The first term on the right side of the equation measures the earnings gap due to characteristics (the "explained" part). The second one corresponds to differences in returns between the two groups, which may come from unobservable characteristics and/or discrimination (the "unexplained" part or residual). Oaxaca and Blinder decompositions have recently become contended as the method requires us to determine a priori which earnings structure is non discriminatory, resulting in the so-called index problem. If we choose the coefficients of the high-earning groups (ethnic majority or males) as non-discriminatory, the low-earning groups (ethnic minority or females) are assumed to be underpaid or facing discriminated. Meanwhile, if the coefficients of the low-earning group are considered as the non-discriminatory benchmark, the other groups are supposed to receive patronage.

We use Neumark's (1988) approach to obtain a non-discriminatory coefficient. Following Neumark's method, we obtain $\beta^*$, a non-discriminatory coefficient, by estimating the pooled sample using the weighted average of the earnings structure of the high earning and low earning groups.

$$\overline{lnW_k} - \overline{lnW_n} = \beta^*(\bar{x}_k - \bar{x}_n) + [(\beta_k - \beta^*)\bar{x}_k + (\beta^* - \beta_n)\bar{x}_n] \quad (6)$$

The first term on the right-hand side is the earnings gap attributable to differences in characteristics while the second and the third terms represent the earning differential between the actual and pooled returns to characteristics for the two groups. While Neumark's approach overcomes the index problem, it is not able to take into account selection problems stemming from sectoral choices of the high earning group and low earning. Appleton et al. (1999) then proposes a full decomposition method which takes into account differences in sectoral choice while following Neumark's approach to avoid the index problem. The full decomposition starts with the idea of considering the ethnic majority's (or male) and minority's (or female) earnings as the sum of sectoral earnings weighted by the proportion of workers of each group in each sector. Let $\bar{p}_{ks}$ and $\bar{p}_{ns}$ be the sample proportion of Kinh (male) and non-Kinh (female) participating in each sector s. Thus, their earnings then can be expressed as:

$$\bar{W}_k = \sum_{s=1}^{4} \bar{W}_{ks}\bar{p}_{ks}$$

$$\bar{W}_n = \sum_{s=1}^{4} \bar{W}_{ns}\bar{p}_{ns}$$

Appleton et al. (1999)'s full decomposition can be written as follows:

$$\bar{W}_k - \bar{W}_n = \sum_{s=1}^{4} \bar{p}_s^* (\bar{W}_{ks} - \bar{W}_{ns}) + \sum_{s=1}^{4} \bar{W}_{ks} (\bar{p}_{ks} - \bar{p}_s^*) + \sum_{s=1}^{4} \bar{W}_{ns}(\bar{p}_s^* - \bar{p}_{ns}) \quad (7)$$

Equation 7 builds on the necessary assumption that the sectoral structure would prevail even in the absence of ethnic (or gender) earning differences that could impact sectoral choice. Under this assumption, $p_s^*$ then represents the proportion of individuals in each sector s. The first term $(\sum_{s=1}^{4} \bar{p}_s^* (\bar{W}_{ks} - \bar{W}_{ns}))$ can be decomposed using the Neumark decomposition. The second and third terms can also be decomposed further to separate the differences arising from differences in observable characteristics and differences attributed to the returns to those characteristics. We could derive the average probability for Kinh (male) and non-Kinh (female) employees in different sectors by estimating a multinominal logit equation separately for Kinh (male) and non-Kinh (female): $\bar{p}_{ks}$ and $\bar{p}_{ns}$. The multinomial logit equation in this section includes four modalities of sectors (s): i) public sector; ii) formal sector; iii) informal sector and iv) agricultural sector. After taking into account selectivity of each group in each sector, the equation (7) can be expanded into seven terms:

$$\bar{W}_k - \bar{W}_n = \sum_{s=1}^{4} \bar{p}_s^* (\overline{x_{ks}} - \overline{x_{ns}})\beta_s^* + \sum_{s=1}^{4} \bar{p}_s^* \bar{x}_{ks}(\beta_{ks} - \beta_s^*) + \sum_{s=1}^{4} \bar{p}_s^* \bar{x}_{ns}(\beta_s^* - \beta_{ns}) + \sum_{s=1}^{4} \bar{W}_{ks} (\bar{p}_{ks}^* - \bar{p}_s^*) + \sum_{s=1}^{4} \bar{W}_{ns}(\bar{p}_s^* - \bar{p}_{ns}^*) + \sum_{s=1}^{4} \bar{W}_{ks} (\bar{p}_{ks} - \bar{p}_{ks}^*) + \sum_{s=1}^{4} \bar{W}_{ns}(\bar{p}_{ns}^* - \bar{p}_{ns}) \quad (8)$$

The first three terms represent Neumark's decomposition of within sector earnings gaps. The fourth and fifth terms capture the earning differences arising from the differences in characteristics determining the differences in their sectoral choice structure. The last two terms measure the earning differences due to the deviations between predicted and actual sectoral compositions of the ethnic majority (men) and ethnic minority (women) not accounted for by differences in characteristics.

# 5 Results

## 5.1 Earnings determinants

### 5.1.1 Mincer equation estimation

Table 5 presents the estimates of the pooled OLS regression following equation (1). The goodness of fit of the model is similar to what is usually obtained in the literature, increasing from model 1 to model 3 (adjusted $R^2$ improving from 0.14 to 0.26). Gender and human capital are highly significant: women suffer a penalty between -15 percent and -17 percent compared to their male counterparts, depending on the specifications. Returns to education are convex as in many other studies in developing countries (Normand et al., 2011). Returns to potential experience are concave but increasing with the professional cycle as the maximum is achieved after more than 30 years of experience. The institutional sectors are ordered according to our expectations: from agriculture – the lowest rewarding sector, followed by the informal sector, the formal private sector and at the top of the payment ladder, the public sector and foreign enterprises. Finally, location matters: province fixed effect are at play, suggesting that local labour markets are not fully integrated.

Regarding ethnicity, in the first sets of models, which consider only the binary partition between Kinh and non-Kinh, the coefficient is highly significant, with non-Kinh receiving a wage penalty between -35 percent (model 1) and -22 percent (model 3) compared to Kinh. At this stage, two provisional conclusions can be drawn: first, it seems that in Vietnam, the ethnic earnings gap matters more than the gender gap; second, there is a strong presumption that ethnic minorities are suffering from "poor" location (concentration in remote areas) and even more importantly, job segregation. Moving from model 1, our simplest model, to model 3 which takes into account institutional sectors and location reduces the ethnic earnings gap by 10 percentage points (from -32 percent to -22 percent).

Due to the importance of the ethnic earnings gap we next disaggregate the non-Kinh category into different ethnic groups to better reflect ethnic diversity. In general, Kinh earn more than their fellow citizens as 42 of the 57 estimated earnings coefficients (for 19 different ethnic groups and 3 specifications) are significantly negative. In the simple Mincerian equation (model 1), only Hoa, Kho-me and Ê Đê have higher earnings, while Ba Na and Chăm gain as much as Kinh do. In the full model (model 3), Ê Đê, Ba Na, Kho-me and the residual Other minorities are in line with Kinh, Co-ho and Chăm are the only minority to gain more. The range coefficients is extremely heterogeneous, suggesting vast differences between ethnic groups. Consistent with previous findings, Hmông are the most disadvantaged group. These results confirm the need to differentiate between the different ethnic minorities for analytical and policy purposes.

**Table 5: Labor earnings models**
(Log of hourly income; OLS)

| | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | **Coef.** | **SE** | **Coef.** | **SE** | **Coef.** | **SE** |
| Female | -0.189*** | 0.004 | -0.163*** | 0.004 | -0.173*** | 0.004 |
| Married | 0.112*** | 0.005 | 0.145*** | 0.005 | 0.186*** | 0.005 |
| Years of schooling | -0.082*** | 0.002 | -0.051*** | 0.002 | -0.053*** | 0.002 |
| (Years of schooling)$^2$ | 0.008*** | 0.000 | 0.007*** | 0.000 | 0.005*** | 0.000 |
| Potential experience | 0.043*** | 0.000 | 0.039*** | 0.000 | 0.041*** | 0.000 |
| (Potential experience)$^2$ | -0.001*** | 0.000 | -0.001*** | 0.000 | -0.001*** | 0.000 |
| **Institutional sector** | | | | | | |
| Foreign enterprise | | | | | 0.033** | 0.014 |
| Domestic enterprise | | | | | -0.055*** | 0.009 |
| Non-farm formal HB | | | | | -0.222*** | 0.007 |
| Informal sector | | | | | -0.395*** | 0.007 |
| Agriculture | | | | | -0.825*** | 0.007 |
| Rural | | | | | 0.035*** | 0.004 |
| Intercept | 0.854*** | 0.011 | 0.625*** | 0.017 | 1.103*** | 0.018 |
| Hoa | 0.135*** | 0.019 | 0.001*** | 0.019 | -0.066*** | 0.018 |
| Thái | -0.389*** | 0.016 | -0.494*** | 0.018 | -0.299*** | 0.018 |
| Tày | -0.507*** | 0.011 | -0.425*** | 0.014 | -0.268*** | 0.013 |
| Muong | -0.557*** | 0.019 | -0.520*** | 0.022 | -0.315*** | 0.021 |
| Kho-me | 0.078*** | 0.019 | -0.051*** | 0.019 | 0.014 | 0.019 |
| Nùng | -0.598*** | 0.017 | -0.599*** | 0.019 | -0.427*** | 0.019 |
| Hmông | -1.430*** | 0.023 | -1.163*** | 0.024 | -0.957*** | 0.023 |
| Dao | -0.938*** | 0.024 | -0.822*** | 0.024 | -0.614*** | 0.024 |
| Ê Đê | 0.265*** | 0.038 | -0.134*** | 0.039 | -0.009 | 0.038 |
| Gia-rai | -0.212*** | 0.038 | -0.484*** | 0.037 | -0.281*** | 0.036 |
| Ba na | 0.025 | 0.036 | -0.228*** | 0.039 | -0.048 | 0.038 |
| Xo-dang | -0.476*** | 0.037 | -0.652*** | 0.037 | -0.475*** | 0.036 |
| Ra-glai | -0.792*** | 0.045 | -0.706*** | 0.044 | -0.507*** | 0.043 |
| Co-ho | 0.267*** | 0.049 | 0.142*** | 0.048 | 0.281*** | 0.046 |
| Chăm | 0.001 | 0.049 | -0.036 | 0.048 | 0.102** | 0.046 |
| Giáy | -0.685*** | 0.044 | -0.641*** | 0.043 | -0.454*** | 0.042 |
| Other Northern | -0.736*** | 0.033 | -0.595*** | 0.032 | -0.449*** | 0.031 |
| Other Central Highland | -0.593*** | 0.028 | -0.668*** | 0.027 | -0.520*** | 0.027 |
| Others | -0.204*** | 0.065 | -0.171*** | 0.063 | -0.010 | 0.061 |
| **Non Kinh** | **-0.437*** ** | **0.006** | **-0.390*** ** | **0.007** | **-0.255*** ** | **0.007** |
| **Province fixed effect** | No | | Yes | | Yes | |
| **No. Observation** | 310,154 | | 310,154 | | 310,154 | |
| **Adjusted $R^2$** | 0.142 | | 0.217 | | 0.260 | |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%. Other coefficients in the aggregated ethnicity model of Kinh/non Kinh equation are not reported.

To investigate further the gender and ethnic earning gap, we restrict our sample to nonfarm labour income (Table 6). Different reasons can be suggested to justify such an option. First, agricultural incomes are among the most difficult to capture through a single question, like in classical labour force surveys. Second, agriculture labour market is known to be functioning very differently from nonfarm sectors one, which implies that earnings functions should be sector specific. Third, as we saw that job segregation is probably at play, and ethnic minorities are

concentrated in farming activities, it is all the more interesting to focuses on what is happening outside agriculture. Finally, with more than 10,000 ethnic minorities observations (whether for nonfarm workers or for wage workers), we able to go beyond previous studies which had to focus only on rural issues.

**Table 6: Non-farm labor earnings models**
(Log of hourly income; OLS)

| | **Model 1** | | **Model 2** | | **Model 3 (only wage earners)** | |
|---|---|---|---|---|---|---|
| | **Coef.** | **SE** | **Coef.** | **SE** | **Coef.** | **SE** |
| Female | -0.169*** | 0.004 | -0.169*** | 0.004 | -0.111*** | 0.003 |
| Married | 0.080*** | 0.005 | 0.095*** | 0.005 | 0.047*** | 0.004 |
| Years of schooling | -0.057*** | 0.002 | -0.046*** | 0.002 | -0.017*** | 0.001 |
| (Years of schooling)$^2$ | 0.006*** | 0.000 | 0.006*** | 0.000 | 0.004*** | 0.000 |
| Potential experience | 0.037*** | 0.000 | 0.039*** | 0.000 | 0.033*** | 0.000 |
| (Potential experience)$^2$ | -0.001*** | 0.000 | -0.001*** | 0.000 | -0.001*** | 0.000 |
| Intercept | 1.112*** | 0.012 | 1.083*** | 0.013 | 1.220*** | 0.009 |
| Hoa | 0.006 | 0.017 | -0.038** | 0.017 | -0.089*** | 0.015 |
| Thái | 0.140*** | 0.032 | -0.006 | 0.032 | 0.027 | 0.021 |
| Tày | 0.119*** | 0.014 | 0.015 | 0.017 | 0.052*** | 0.012 |
| Muong | -0.029 | 0.032 | -0.026 | 0.033 | -0.083*** | 0.021 |
| Kho-me | -0.031 | 0.024 | -0.056** | 0.024 | -0.053*** | 0.014 |
| Nùng | 0.122*** | 0.025 | -0.041 | 0.027 | -0.025 | 0.020 |
| Hmông | -0.109 | 0.104 | -0.273*** | 0.102 | -0.044 | 0.062 |
| Dao | 0.106* | 0.064 | -0.013 | 0.063 | -0.061 | 0.040 |
| Ê Đê | 0.010 | 0.096 | -0.184* | 0.095 | 0.119*** | 0.040 |
| Gia-rai | 0.115 | 0.118 | -0.128 | 0.116 | -0.084** | 0.042 |
| Ba na | -0.008 | 0.111 | -0.231** | 0.106 | -0.126** | 0.052 |
| Xo-dang | 0.241 | 0.158 | 0.073 | 0.151 | -0.000 | 0.070 |
| Ra-glai | -0.429*** | 0.150 | -0.360** | 0.147 | 0.068 | 0.068 |
| Co-ho | -0.183 | 0.213 | -0.162 | 0.208 | -0.117 | 0.078 |
| Chăm | -0.160** | 0.080 | -0.144* | 0.078 | 0.003 | 0.045 |
| Giáy | 0.187** | 0.087 | 0.031 | 0.086 | -0.049 | 0.058 |
| Other Northern | 0.029 | 0.055 | -0.034 | 0.054 | -0.033 | 0.034 |
| Other Central Highland | 0.100 | 0.074 | -0.014 | 0.072 | -0.129*** | 0.033 |
| Other | 0.250* | 0.128 | 0.200 | 0.125 | 0.284*** | 0.071 |
| **Non Kinh** | **-0.061*** ** | **0.008** | **-0.029*** ** | **0.009** | **-0.028*** ** | **0.006** |
| **Province fixed effect** | No | | Yes | | Yes | |
| **No. Observation** | 200,413 | | 200,413 | | 117,253 | |
| **Adjusted $R^2$** | 0.125 | | 0.217 | | 0.367 | |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%. Other coefficients in the aggregated ethnicity model of Kinh/non Kinh equation are not reported.

Obviously, running regressions on nonfarm paid workers raises an issue of selection bias: ethnic minorities working outside agriculture is not a random subsample, and the unobserved factors that pushed them out of agriculture are probably also impacting positively their earnings. Aware of this limitation, as an initial step, we consider nevertheless that our "naive" estimations may be informative. Not to add new endogeneity problems to the previous issue, we estimate our simplest mincerian model, with and without province fixed effect (model 2 and 1). Model 3 considers only wage workers as wages are even better measured through the LFS (than profits) and as the potential discrimination of employees by their employers is better assessed than for non-wage workers, where discrimination can only occurs through the product markets, as they are their own employers.

Findings from Table 6 show that nonfarm ethnic gap is highly reduced. At the aggregate level, the gap, though still significant, ranges from -6 percent (model 1) to -3 percent (models 2 and 3). Compared to the gender gap, the ethnic gap if negligible: gender gap remains at nearly 16 percent against women for nonfarm jobs, and 10.5 percent for female wage workers. Once the ethnic groups disaggregated, many of them earns as much or even more than Kinh. Only 13 of the 57 estimated coefficient are significant at 1 percent level. The highest penalty is suffered by Ra-glai (-0.42*** in model 1 and -0.36*** in model 2) for nonfarm jobs, and a few Central Highland minorities (Ê Đê, Ba na and Other).

The interpretation of these results is not straightforward. Taken at face value, our estimates suggest that potential earning "discrimination" against ethnic minorities on the labour market is very low, once they get out of farm job or when they work as employees. The whole question resides in whether those who succeed in this process are selected (or self-selected) according to some specific skills outside our human capital indicators (ability, for instance) or not. In econometric terms as quoted above, is there any unobservable variable playing on both sectoral allocation and labour income? In order to address this selection bias, we will use the multinomial logit model to correct for sectoral allocation in the following part.

### 5.1.2 Mincer equation estimation with selectivity correction

In this section, our results are reported in two parts. The first part (Tables 7a, b) depicts marginal effects from the multinomial logit model and second part (Tables 8a, b) highlights the results from the earnings equations estimation, in which the selection bias is corrected by a multinomial logit model.

The marginal effects from the multinomial logit estimation/ the first stage equations show us the determinants of the allocation of people into different sectors. As expected, we observe a positive effect of education for public sector, formal, and informal sector employment. The effect is strongest in the public sector, convex in formal private sector and concave in informal sector while it has negative effect on allocation into agriculture. The convex influence of education on selection into formal sector could be explained by the fact that there is a range of formal jobs available that does not require high education, such as workers in factories and non-farm formal household business. As expected, experience also has a positive marginal effect on selection into different sectors and a negative effect on being inactive. People coming from larger households have a lower probability of working in the public sector, the informal sector, and agriculture sector while they have higher probability of working in the formal sector and

being inactive. It is also seen that people are less likely to work in public and formal sectors when having high dependence ratio. The opposite effect is observed in informal and agriculture sector.

**Table 7a: Determinants of the allocation of people to different employment sectors**
(Marginal effect from multinomial logit model on sectoral choice)

| | (1) Public | (2) Formal | (3) Informal | (4) Agriculture | (5) Inactive |
|---|---|---|---|---|---|
| Female | 0.0027*** | -0.0214*** | -0.0195*** | -0.0583*** | 0.0965*** |
| | (0.0005) | (0.0016) | (0.0017) | (0.0018) | (0.0022) |
| Married | 0.0189*** | 0.0220*** | 0.0343*** | 0.0811*** | -0.156*** |
| | (0.0008) | (0.0018) | (0.0023) | (0.0025) | (0.0030) |
| Non-Kinh | 0.009*** | -0.0486*** | -0.114*** | 0.156*** | -0.0025 |
| | (0.0010) | (0.0027) | (0.0036) | (0.0027) | (0.0042) |
| Years of schooling | 0.0060*** | -0.0052*** | 0.0107*** | -0.0061*** | -0.0054*** |
| | (0.0004) | (0.0007) | (0.0009) | (0.0008) | (0.0011) |
| (Years of schooling)$^2$ | 0.0006*** | 0.0010*** | -0.0018*** | -0.0015*** | 0.0016*** |
| | (2.30e-05) | (3.55e-05) | (5.20e-05) | (5.20e-05) | (6.16e-05) |
| Potential experience | 0.0052*** | 0.0117*** | 0.0177*** | 0.0116*** | -0.0463*** |
| | (7.92e-05) | (0.0002) | (0.0002) | (0.0002) | (0.0003) |
| (Potential experience)$^2$ | -0.0001*** | -0.0003*** | -0.0003*** | -0.0002*** | 0.0009*** |
| | (1.64e-06) | (2.75e-06) | (3.24e-06) | (2.87e-06) | (4.07e-06) |
| Rural | -0.0116*** | -0.0756*** | -0.0412*** | 0.291*** | -0.162*** |
| | (0.0006) | (0.0014) | (0.0017) | (0.0017) | (0.0023) |
| Spouse of the Head | -0.0114*** | -0.0383*** | -0.0440*** | -0.0053** | 0.0991*** |
| | (0.0007) | (0.0017) | (0.0022) | (0.0021) | (0.0032) |
| Children of the Head | -0.0021** | -0.0074*** | -0.0045 | -0.0918*** | 0.106*** |
| | (0.0009) | (0.0022) | (0.0030) | (0.0033) | (0.0045) |
| Parents of the Head | -0.0152** | -0.0377** | 0.0063 | -0.185*** | 0.231*** |
| | (0.0068) | (0.0158) | (0.0121) | (0.0097) | (0.0128) |
| Familiy ties | -0.0006 | -0.0061** | -0.0038 | -0.0802*** | 0.0907*** |
| | (0.0011) | (0.0026) | (0.0035) | (0.0038) | (0.0049) |
| Non-family ties | -0.0020 | 0.112*** | 0.126*** | -0.0920*** | -0.144*** |
| | (0.0056) | (0.0097) | (0.0160) | (0.0243) | (0.0238) |
| Dependence ratio | -0.0153*** | -0.0379*** | 0.132*** | 0.0653*** | -0.144*** |
| | (0.0015) | (0.0036) | (0.0046) | (0.0045) | (0.0067) |
| HH size | -0.0013*** | 0.0027*** | -0.0067*** | -0.0015*** | 0.0067*** |
| | (0.0002) | (0.0004) | (0.0005) | (0.0005) | (0.0007) |
| Province fixed effect | Yes | Yes | Yes | Yes | Yes |
| Observations | 453667 | 453667 | 453667 | 453667 | 453667 |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

**Table 7b: Determinants of the allocation of people to different employment sectors**

(Marginal effect from multinomial logit model on sectoral choice, disaggregated ethnic groups)

| | (1) Public | (2) Formal | (3) Informal | (4) Agriculture | (5) Not working |
|---|---|---|---|---|---|
| Hoa | -0.0155* | 0.0736* | -0.000268 | -0.0920* | 0.0341* |
| | (0.00847) | (0.0380) | (0.00661) | (0.0483) | (0.0196) |
| Thái | 0.0154 | -0.0906 | -0.193 | 0.214*** | 0.0543 |
| | (0.0122) | (0.111) | (0.272) | (0.0154) | (0.164) |
| Tày | 0.0159*** | -0.105 | -0.134 | 0.182*** | 0.0410 |
| | (0.00585) | (0.0984) | (0.190) | (0.0269) | (0.112) |
| Muong | 0.0119 | -0.0885 | -0.136 | 0.183*** | 0.0296 |
| | (0.00845) | (0.0911) | (0.193) | (0.0273) | (0.121) |
| Kho-me | 0.00501 | -0.0173 | -0.0238 | 0.109** | -0.0733 |
| | (0.00370) | (0.0181) | (0.0343) | (0.0448) | (0.0622) |
| Nùng | 0.00992* | -0.106 | -0.0930 | 0.191*** | -0.00204 |
| | (0.00517) | (0.0854) | (0.132) | (0.0520) | (0.0940) |
| Hmông | 0.0412** | -0.223 | -0.356 | 0.373*** | 0.165 |
| | (0.0174) | (0.235) | (0.505) | (0.0220) | (0.270) |
| Dao | 0.0165 | -0.191 | -0.180 | 0.287*** | 0.0671 |
| | (0.0114) | (0.159) | (0.254) | (0.0586) | (0.145) |
| Ê Đê | 0.0246 | -0.106 | -0.287 | 0.229*** | 0.139 |
| | (0.0190) | (0.151) | (0.406) | (0.0305) | (0.214) |
| Gia-rai | 0.0763*** | -0.416 | -0.284 | 0.447*** | 0.177 |
| | (0.0143) | (0.329) | (0.405) | (0.0932) | (0.202) |
| Ba na | 0.0195 | -0.104 | -0.221 | 0.289*** | 0.0163 |
| | (0.0163) | (0.130) | (0.313) | (0.0418) | (0.213) |
| Xo-dang | 0.197 | 0.259 | -3.076 | 0.979 | 1.641 |
| | (9.558) | (30.57) | (179.2) | (46.83) | (92.21) |
| Ra-glai | 0.0451*** | -0.265 | -0.234 | 0.318*** | 0.135 |
| | (0.0139) | (0.224) | (0.333) | (0.0523) | (0.169) |
| Co-ho | 0.0495*** | -0.338 | -0.273 | 0.231*** | 0.330*** |
| | (0.0177) | (0.289) | (0.392) | (0.0406) | (0.124) |
| Chăm | 0.00707 | -0.130 | -0.0991 | 0.171*** | 0.0508 |
| | (0.0113) | (0.103) | (0.142) | (0.0425) | (0.0786) |
| Giáy | -0.00277 | -0.114 | -0.0818 | 0.183*** | 0.0147 |
| | (0.0129) | (0.0899) | (0.119) | (0.0561) | (0.0813) |
| OtherNorthern minorities | 0.0219* | -0.109* | -0.0133 | 0.145** | -0.0451 |
| | (0.0113) | (0.0637) | (0.0268) | (0.0694) | (0.0449) |
| Other Central Highland | 0.0587*** | -0.253 | -0.374 | 0.291*** | 0.277 |
| | (0.0100) | (0.256) | (0.529) | (0.0411) | (0.230) |
| Other | 0.0471** | -0.0708 | -0.388 | 0.268*** | 0.144 |
| | (0.0189) | (0.169) | (0.552) | (0.0638) | (0.315) |
| Control variables | Yes | Yes | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes | Yes | Yes |
| Observations | 453667 | 453667 | 453667 | 453667 | 453667 |

Source: LFS2007, GSO; authors' calculations.

Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

Regarding our variable of interest, women are more likely to work in the public sector but, ceteris paribus, face a higher probability than men of not working at all. Ethnic minorities have a higher probability of working in the public sector and in agriculture, while they have lower probability of working in the formal or informal private sector. Other Central Highland groups and other groups have higher probability of working in the public sector compared to Kinh, Tày, Nùng, Hmông, Gia Rai, Ra-glai, Co-ho, and other Northern minorities, while Hoa are less likely to work in the public sector than Kinh. Regarding formal and informal private sectors, there is almost no significant difference in the allocation among different ethnic groups. Lastly, almost all ethnic minorities groups have a higher probability of engaging in agriculture than Kinh (except for Hoa).

*Earning equation estimates*

After correcting for selectivity into the four different sectors, Table 8a and Table 8b present the earning equation estimations. It is seen that returns to education are positive and significant in all cases but the association is convex in the public sector and concave in the informal and agriculture sector. The potential experience and earnings relationship is concave in public and informal sector but convex in the other sectors.

Women have slightly higher earnings than men in the public and agricultural sector after accounting for selection, by 1.5 percent and 4.5 percent, respectively. However, all else equal, women face an earning gap of 11.5 percent and 18 percent less in the formal and informal sectors, compared to men. Women get higher earnings in agriculture than men could be explained by two reasons. First, our classification of sector does not take into account employment status. Thus, wage work in agriculture could drive the result. Second, it is possible that women have more interrupted work periods than men, thus, our estimation could be biased in the sense that we overestimate the experience effects for women. However, the LFS does not have information on labor market history which could have been used to control for this potential bias. Meanwhile, the ethnic earnings gaps vary considerably between different sectors. As an ethnic minority, an individual can get an earning premium of up to 2 percent than their Kinh counterparts if they work in public sector, however, their earning will be 14 percent and 16 percent less when they are engaged in formal and informal employment. Moreover, we find a large gap in earnings between Kinh and non-Kinh in the agricultural sector, which is 45 percent. This result is in line with Baulch and Vu (2011)'s finding that the ethnic expenditure gap in rural area is 58.1 percent in 2004 and 68.1 percent in 2010. Even after correcting for selection into different sectors, we still find a large gap in agriculture. It is likely that Kinh in agriculture might be doing for profit large scale while non-Kinh are engaged in small scale subsistence agriculture, which could not be covered in the selection equations.

To explore the heterogeneity of effects, we interact the dummies of gender and ethnicity for each regression. It is seen that in public sector, once the interaction is introduced, the positive and significant effect of being a women or being an ethnic minority does not hold. Only women in ethnic minority groups benefit higher earning in public sector. In the last column of Table 8a, we find that the significant and positive effect of gender dummy was captured by the effect of woman in ethnic minority group. There is indeed no significant difference in earning between Kinh male and Kinh female workers in agriculture. In short, female ethnic minorities have vastly

lower earnings than Kinh or non-Kinh men in all but public and barely in agriculture sectors.

**Table 8a: Earning equation estimates for different employment sectors**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | **Public** | | **Formal** | | **Informal** | | **Agriculture** | |
| Female | 0.015*** | 0.007 | -0.124*** | -0.127*** | -0.202*** | -0.205*** | 0.0444*** | 0.004 |
| | (0.005) | (0.006) | (0.007) | (0.007) | (0.01) | (0.012) | (0.0066) | (0.014) |
| Ethnic minority | 0.025* | -0.020* | -0.150*** | -0.174*** | -0.177*** | -0.206*** | -0.593*** | -0.650*** |
| | (0.014) | (0.011) | (0.019) | (0.040) | (0.027) | (0.011) | (0.038) | (0.045) |
| Female x Ethnicity | | 0.089*** | | 0.054 | | 0.056* | | 0.128*** |
| | | (0.015) | | (0.057) | | (0.030) | | (0.008) |
| Married | -0.003 | -0.001 | -0.187*** | -0.187*** | -0.036** | -0.036*** | -0.089*** | -0.083*** |
| | (0.005) | (0.013) | (0.013) | (0.017) | (0.015) | (0.005) | (0.018) | (0.018) |
| Years of schooling | -0.076*** | -0.075*** | 0.033*** | 0.033*** | 0.038*** | 0.039*** | 0.061*** | 0.064*** |
| | (0.008) | (0.011) | (0.003) | (0.006) | (0.004) | (0.007) | (0.004) | (0.004) |
| (Years of schooling)$^2$ | 0.004*** | 0.0041*** | 0.000 | 0.000 | -0.003*** | -0.003*** | -0.002*** | -0.002*** |
| | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Potential experience | 0.028*** | 0.028*** | -0.035*** | -0.035*** | 0.010*** | 0.010*** | -0.008*** | -0.006* |
| | (0.001) | (0.002) | (0.001) | (0.005) | (0.002) | (0.001) | (0.0018) | (0.003) |
| (Potential experience)$^2$ | -0.000*** | -0.000*** | 0.001*** | 0.001*** | -0.000** | -0.000*** | 0.0005*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Rural area | 0.064*** | 0.063*** | -0.080*** | -0.081*** | -0.094*** | -0.095*** | -0.691*** | -0.679*** |
| | (0.012) | (0.008) | (0.024) | (0.010) | (0.016) | | (0.0389) | (0.042) |
| BFG Public | -0.140*** | -0.135*** | -0.914*** | -0.916*** | -0.915*** | -0.918*** | -1.053*** | -1.137*** |
| | (0.013) | (0.022) | (0.077) | (0.073) | (0.060) | (0.065) | (0.102) | (0.079) |
| BFG Formal | -0.429*** | -0.436*** | -0.075*** | -0.075*** | 0.146*** | 0.151*** | 0.962*** | 1.054*** |
| | (0.117) | (0.149) | (0.021) | (0.028) | (0.0566) | (0.027) | (0.110) | (0.082) |
| BFG Informal | 0.432*** | 0.442*** | -0.441*** | -0.440*** | 0.0167 | 0.017** | 0.713*** | 0.675*** |
| | (0.046) | (0.0251) | (0.079) | (0.083) | (0.0116) | (0.008) | (0.163) | (0.135) |
| BFG Agriculture | 0.164*** | 0.153** | -0.783*** | -0.786*** | -0.315*** | -0.317*** | -0.163*** | -0.156*** |
| | (0.020) | (0.0637) | (0.088) | (0.056) | (0.0458) | (0.042) | (0.013) | (0.026) |
| BFG Inactive | 0.104*** | 0.0979 | 1.032*** | 1.030*** | 0.658*** | 0.656*** | 2.306*** | 2.256*** |
| | (0.029) | (0.0662) | (0.069) | (0.062) | (0.0670) | (0.028) | (0.049) | (0.104) |
| Constant | 2.22*** | 2.199*** | 1.846*** | 1.845*** | 1.292*** | 1.292*** | 1.800*** | 1.763*** |
| | (0.162) | (0.177) | (0.143) | (0.175) | (0.0656) | (0.059) | (0.098) | (0.216) |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%. The figures in parentheses are bootstrapped standard errors.

**Table 8b: Earning equation estimates for different employment sectors**
(Disaggregated ethnic groups)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | **Public** | **Formal** | **Informal** | **Agriculture** |
| Hoa | -0.0305 | 0.0444 | 0.0409 | 0.246*** |
| | (0.0387) | (0.0414) | (0.0442) | (0.0813) |
| Thái | 0.0618*** | -0.132* | -0.235*** | -0.592*** |
| | (0.0185) | (0.0728) | (0.0205) | (0.0456) |
| Tày | 0.0866*** | -0.229*** | -0.347*** | -0.459*** |
| | (0.0239) | (0.0710) | (0.0452) | (0.0542) |
| Muong | -0.0965* | -0.115** | -0.113*** | -0.614*** |
| | (0.0514) | (0.0502) | (0.0374) | (0.0337) |
| Kho-me | -0.0540 | -0.192*** | -0.103*** | -0.312*** |
| | (0.0344) | (0.0289) | (0.0116) | (0.0242) |
| Nùng | 0.0901** | -0.208*** | -0.249*** | -0.578*** |
| | (0.0386) | (0.0519) | (0.0676) | (0.0433) |
| Hmông | -0.00738 | -0.884 | -0.809 | -0.972*** |
| | (0.0785) | (0.696) | (0.649) | (0.0503) |
| Dao | 0.0837 | -0.0577 | -0.131** | -0.740*** |
| | (0.0750) | (0.0899) | (0.0633) | (0.0746) |
| Ê Đê | -0.184 | -0.00524 | -0.442*** | -0.581*** |
| | (0.165) | (0.0455) | (0.107) | (0.0508) |
| Gia-rai | -0.0551 | -0.488** | -0.494*** | -1.053*** |
| | (0.113) | (0.246) | (0.0576) | (0.0665) |
| Ba na | -0.156 | -0.702** | -0.101* | -0.611*** |
| | (0.130) | (0.334) | (0.0578) | (0.0474) |
| Xo-dang | 0.0456 | 0.0514 | 0 | -0.923*** |
| | (0.0942) | (0.124) | (0) | (0.0301) |
| Ra-glai | -0.191* | 0.185 | -0.614 | -1.052*** |
| | (0.0997) | (0.159) | (0.380) | (0.109) |
| Co-ho | 0.0594 | -0.278*** | -0.467** | -0.241*** |
| | (0.192) | (0.0872) | (0.235) | (0.0805) |
| Chăm | 0.110 | -0.132 | -0.311*** | -0.238*** |
| | (0.0868) | (0.0884) | (0.0878) | (0.0555) |
| Giáy | 0.189*** | 0.0956 | -0.0511 | -0.655*** |
| | (0.0393) | (0.140) | (0.0610) | (0.123) |
| Other Northern minorities | 0.0439*** | -0.325 | -0.170** | -0.593*** |
| | (0.0112) | (0.278) | (0.0670) | (0.0491) |
| Other Central Highland | 0.0593 | -0.136 | -0.722*** | -0.887*** |
| | (0.0653) | (0.0962) | (0.257) | (0.0910) |
| Others | 0.255*** | -0.176*** | -0.289*** | -0.361*** |
| | (0.0467) | (0.0658) | (0.0931) | (0.130) |
| BFG_public | -0.139*** | -0.935*** | -0.900*** | -0.644*** |
| | (0.0109) | (0.0891) | (0.0633) | (0.0940) |
| BFG_Formal | -0.432*** | -0.0721*** | 0.173*** | 0.439*** |
| | (0.116) | (0.0230) | (0.0617) | (0.137) |
| BFG_Informal | 0.431*** | -0.440*** | 0.0157 | 0.567*** |

| | | | | |
|---|---|---|---|---|
| | (0.0474) | (0.0860) | (0.0107) | (0.177) |
| BFG_Agriculture | 0.170*** | -0.786*** | -0.295*** | -0.105*** |
| | (0.0235) | (0.0972) | (0.0616) | (0.0152) |
| BFG_Inactive | 0.104*** | 1.048*** | 0.686*** | 2.274*** |
| | (0.0311) | (0.0865) | (0.0669) | (0.0478) |
| Constant | 2.230*** | 1.859*** | 1.334*** | 1.378*** |
| | (0.147) | (0.161) | (0.0628) | (0.112) |
| Control variables | Yes | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes | Yes |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%. The figures in parentheses are bootstrapped standard errors.

Estimating the earnings equations with disaggregated ethnic groups depicts once again a very heterogeneous coefficient between ethnic minority groups. In public sector, Thái, Tày, Nùng, Giáy, other Northern minorities and the other group earn more than Kinh while Muong and Ra-glai earn less. In the formal sector, the coefficient is negative and significant for almost all ethnic minority groups, mostly around -0.2 but in the extreme case, we find a coefficient of -0.7 for Ba Na group which is significant at 5 percent level. In the informal sector, we also find very large range in the coefficients, from -0.1 (Muong, Kho-me, Ba Na) to -0.7 (Other Central Highland minorities). In line with our discussion above, the ethnic earning gap is becoming significantly large when people work in agriculture sector and the gap varies from the lowest of -0.2 (Chăm, Co-ho) to approximately -1.0 (Hmông, Gia Rai, Ra-glai). Through all different sectors, it is notable that Hmông, Gia Rai and Ra-glai are the most disadvantaged ethnic groups. Hoa group, on the contrary, has higher earnings *ceteris paribus* than the majority group, demonstrating again that combining Hoa with Kinh as is done in most academic papers hides important differences between the two groups.

The selectivity correction terms further give us information to understand the selectivity pattern based on unobserved characteristics. In the public sector earning equation, we find a negative selectivity correction term of the formal sector selection equation. This indicates that lower earning of individuals in public sector, compared to individuals taken at random, are due to the allocation of people with worse unobserved characteristic out of the formal private sector and into the public sector. Similarly, we see that the coefficient related to the informal sector selection equation is negative and significant in the formal sector earnings equation. This means that people who would have performed better in the informal sector, on the basis of their unobserved skills, end up working in the formal private sector.

While the earnings regressions presented so far take into account selection into different sectors, they still only provide estimates for earning gaps at the mean of the distribution, thus assuming that differences in terms of the impact of the exogenous variables along the conditional distribution are unimportant. However, one can well imagine that gender and ethnic earnings gaps might vary across the outcome distribution, especially at the lower and higher tails of the earning distribution. We thus estimate quantile earnings regression which allows us to quantify the influence of the different explanatory variables on conditional earnings at the bottom, median and the top of the distribution. Due to the identification issues with running quantile regression with multinomial logit in the first step, we use Heckman procedure to partially correct for selection bias. We use the same set of instrumental variables as in the multinomial logit

model presented in section 4.1 above and estimate the earnings regressions at five points of the log earning distribution, namely 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles.

Results from Table 9 show that the gender earnings gaps do not vary much across the different quantiles while the ethnic earnings gaps are higher at the lower quantiles of the distribution and get narrower at the higher level of the earnings distribution. Besides, the heterogeneity for ethnic minorities becomes clear when we estimate the gaps at different points of the earning distribution. For instance, Hoa people suffer a penalty of 12.4 percent compared to Kinh for participating in a low-paying job (at the $10^{th}$ quantile), but would enjoy a wage premium of 7.2 percent for being involved in a high-paying job ($90^{th}$ quantile). Among ethnic groups which suffer most in terms of earning gaps, such as Hmông, Gia rai, Ra-glai, Nùng, Central Highland, the coefficients are higher at the first two deciles of the distributions which is in line with other studies in developing countries where the earning gap is highest at lowest quantiles of the earning distribution.

At this stage, these results confirm our presumption that in Vietnam, the ethnic earning gap and especially the earnings gap among heterogeneous ethnic minority matters more than the gender earning gap. However, when we look at the earnings gap in non-farm work and for wage earners only, the ethnic gap reduces significantly. This motivates us to further decompose the gaps according to explained and unexplained variation (the latter being commonly understood as discrimination) which is presented in more details in the following section.

**Table 9: Labor earnings quantile regression model**

| | 10th | 25th | 50th | 75th | 90th |
|---|---|---|---|---|---|
| Female | -0.100*** | -0.132*** | -0.143*** | -0.138*** | -0.128*** |
| | (0.0042) | (0.0029) | (0.0025) | (0.0028) | (0.0038) |
| Years of schooling | -0.043*** | -0.0343*** | -0.0242*** | -0.0089*** | 0.00681*** |
| | (0.0022) | (0.0015) | (0.0012) | (0.0014) | (0.0018) |
| (Years of schooling)2 | 0.0045*** | 0.0040*** | 0.0035*** | 0.0029*** | 0.0022*** |
| | (0.0001) | (7.80e-05) | (6.54e-05) | (7.41e-05) | (0.0001) |
| Potential experience | 0.0088*** | 0.0091*** | 0.0100*** | 0.0091*** | 0.0063*** |
| | (0.0010) | (0.0007) | (0.0006) | (0.0010) | (0.0009) |
| (Potential experience)2 | -0.0001*** | -0.0001*** | -0.0001*** | -4.78e-05*** | 5.50e-05*** |
| | (1.81e-05) | (1.26e-05) | (1.09e-05) | (1.24e-05) | (1.66e-05) |
| Institutional sector | | | | | |
| Foreign enterprise | 0.124*** | -0.0278** | -0.136*** | -0.191*** | -0.161*** |
| | (0.0163) | (0.0109) | (0.0093) | (0.0108) | (0.0149) |
| Domestic enterprise | -0.0181* | -0.0953*** | -0.151*** | -0.158*** | -0.103*** |
| | (0.0100) | (0.0066) | (0.0056) | (0.0064) | (0.0089) |
| Non farm formal HB | -0.193*** | -0.157*** | -0.120*** | -0.0490*** | 0.0548*** |
| | (0.0089) | (0.0058) | (0.0048) | (0.0055) | (0.0075) |
| Informal sector | -0.352*** | -0.329*** | -0.322*** | -0.299*** | -0.245*** |
| | (0.0083) | (0.0052) | (0.0043) | (0.0049) | (0.0067) |
| Agriculture | -0.781*** | -0.752*** | -0.679*** | -0.570*** | -0.427*** |
| | (0.0088) | (0.0055) | (0.0045) | (0.00516) | (0.0072) |
| Rural | -0.0403*** | -0.0141*** | -0.0038 | 0.0007 | -0.0069 |
| | (0.0050) | (0.0034) | (0.0030) | (0.00340) | (0.0046) |
| Intercept | 1.298*** | 1.624*** | 1.915*** | 2.152*** | 2.401*** |

| | | | | | |
|---|---|---|---|---|---|
| | (0.0255) | (0.0173) | (0.0147) | (0.0165) | (0.0221) |
| Mills ratio | -0.423*** | -0.319*** | -0.279*** | -0.297*** | -0.346*** |
| | (0.0142) | (0.0099) | (0.0086) | (0.0098) | (0.0132) |
| Hoa | -0.117*** | -0.0580*** | -0.0286** | 0.0340** | 0.0724*** |
| | (0.0201) | (0.0137) | (0.0118) | (0.0134) | (0.0181) |
| Thái | -0.192*** | -0.289*** | -0.311*** | -0.317*** | -0.292*** |
| | (0.0199) | (0.0135) | (0.0116) | (0.0134) | (0.0187) |
| Tày | -0.280*** | -0.254*** | -0.268*** | -0.220*** | -0.162*** |
| | (0.0142) | (0.0099) | (0.0087) | (0.0101) | (0.0135) |
| Muong | -0.284*** | -0.294*** | -0.353*** | -0.317*** | -0.268*** |
| | (0.0236) | (0.0161) | (0.0140) | (0.0164) | (0.0225) |
| Kho-me | 0.00614 | -0.0085 | -0.0232* | -0.0422*** | -0.0619*** |
| | (0.0207) | (0.0141) | (0.0121) | (0.0137) | (0.0186) |
| Nùng | -0.432*** | -0.449*** | -0.411*** | -0.275*** | -0.208*** |
| | (0.0202) | (0.0138) | (0.0122) | (0.0141) | (0.0191) |
| Hmông | -0.603*** | -0.778*** | -0.683*** | -0.534*** | -0.462*** |
| | (0.0254) | (0.0172) | (0.0151) | (0.0177) | (0.0246) |
| Dao | -0.526*** | -0.429*** | -0.449*** | -0.481*** | -0.471*** |
| | (0.0265) | (0.0179) | (0.0154) | (0.0177) | (0.0241) |
| Ê Đê | -0.179*** | 0.0778*** | 0.181*** | 0.102*** | -0.0402 |
| | (0.0413) | (0.0283) | (0.0245) | (0.0277) | (0.0372) |
| Gia-rai | -0.270*** | -0.305*** | -0.274*** | -0.131*** | -0.116*** |
| | (0.0400) | (0.0270) | (0.0234) | (0.0265) | (0.0357) |
| Ba na | 0.0520 | 0.0515* | 0.0517** | -0.0289 | -0.172*** |
| | (0.0417) | (0.0285) | (0.0247) | (0.0281) | (0.0379) |
| Xo-dang | -0.220*** | -0.206*** | -0.345*** | -0.421*** | -0.518*** |
| | (0.0408) | (0.0274) | (0.0235) | (0.0267) | (0.0364) |
| Ra-glai | -0.461*** | -0.394*** | -0.339*** | -0.278*** | -0.195*** |
| | (0.0479) | (0.0321) | (0.0277) | (0.0319) | (0.0437) |
| Co-ho | 0.148*** | 0.0875** | 0.266*** | 0.285*** | 0.513*** |
| | (0.0529) | (0.0356) | (0.0304) | (0.0345) | (0.0464) |
| Chăm | 0.128** | 0.0707** | 0.0775*** | 0.140*** | 0.0749 |
| | (0.0509) | (0.0347) | (0.0300) | (0.0341) | (0.0457) |
| Giáy | -0.241*** | -0.332*** | -0.300*** | -0.383*** | -0.320*** |
| | (0.0468) | (0.0316) | (0.0274) | (0.0313) | (0.0427) |
| OtherNorthern | -0.388*** | -0.380*** | -0.323*** | -0.248*** | -0.230*** |
| | (0.0348) | (0.0236) | (0.0203) | (0.0233) | (0.0319) |
| Other Central Highland | -0.524*** | -0.344*** | -0.310*** | -0.292*** | -0.231*** |
| | (0.0292) | (0.0200) | (0.0174) | (0.0200) | (0.0273) |
| Other | 0.0092 | -0.0153 | -0.0661* | -0.0693 | 0.0296 |
| | (0.0663) | (0.0459) | (0.0395) | (0.0449) | (0.0600) |
| **Non-Kinh** | -0.242*** | -0.231*** | -0.225*** | -0.197*** | -0.172*** |
| | (0.0102) | (0.0041) | (0.0042) | (0.0052) | (0.0095) |
| **Province fixed effect** | | | Yes | | |
| **No. Observation** | | | 310,154 | | |

Source: LFS2007, GSO; authors' calculations.

Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%. Other coefficients in the aggregated ethnicity model of Kinh/non Kinh equation are not reported.

## 5.2 Gender and ethnic earnings gaps decomposition

*Neumark decomposition of gender and ethnic earnings gaps*

In this section, we decompose the gender and ethic earnings gaps found above using Neumark's approach. Similar to what we have seen in the last section, the raw gender earning gap is relatively large (0.224) but not as high as the ethnic earning gap (0.526) (Table 10). When the gaps are decomposed, we find that the differences in individual characteristics such as years of education, work experience, and geographic variables can explain only 15.6 percent of the gender earning gap, while 84.4 percent of the gap are left unexplained. For the ethnic earnings gap, the differences in endowments accounts for nearly 61 percent of the gap, which is expected due to the huge differences in educational endowment and location of residence between the Kinh and non-Kinh groups.

**Table 10: Neumark decomposition of gender and ethnic earnings gaps**

| | Raw earning gap | Explained | Unexplained | % Unexplained |
|---|---|---|---|---|
| Gender earnings gaps | 0.224*** | 0.035 | 0.188 | 84.4 |
| Ethnic earnings gaps | 0.526*** | 0.320 | 0.205 | 39.07 |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

Regarding ethnic earning gaps, not only do we decompose the gap between Kinh and non-Kinh but we also extend our analysis by decomposing the gap between different ethnic minority groups. This is motivated by the result above that show that the earning gaps are very heterogenous once we disaggregate the ethnic minority groups. It thus suggests that some ethnic groups might be more heavily discriminated against than others. We decompose the earning gaps by pair, thus each ethnic group along the row (horizontal axis) is compared to another group along the column (vertical axis). A positive coefficient in this comparison means that the group on the row has higher earnings than the group on the column. Due to the high number of ethnic groups but limited observations in smaller sized groups, we restrict the within minority group analysis to the 10 largest ethnic minorities and the Kinh majority group.

Once we disaggregate the ethnic minority groups, the gaps and the proportion of the unexplained part vary remarkably as presented in Table 11. Decomposing the earning gaps between Kinh and disaggregated ethnic minority groups, shows that Hoa and Ede perform better than Kinh and endowments can explain about 62 percent and 40 percent of the gap, respectively. For 3 of the other large ethnic minority groups (Tày, Thái, Muong), the gaps between them and Kinh are quite similar at around 0.5, half of which can be explained by endowments. The earnings gap between Kho-me and Kinh is much narrow, at 0.079, while the gap between Kinh and Hmông is significantly larger at 1.5. In fact, Hmông lag behind all other groups, partially due to their very low level of (Phung et al., 2014). Meanwhile, Ê Đê has the highest earnings among all groups. The matrix illustrates that no common pattern of ethnic earnings gaps between each ethnic minority group and the ethnic majority group exists. In addition, the unexplained part in the gaps fluctuates from almost zero to 117 percent. Saying that

ethnic earnings between the minority and majority groups exists would thus not treat the data fairly. Instead, ethnic earnings gaps need to be considered on a case by case basis.

**Table 11: Neumark decomposition of ethnic earning gaps between different ethnic groups**

| | Kinh | Hoa | Thái | Tày | Mường | Khơ-me | Nùng | Hmong | Dao | Ê Đê |
|---|---|---|---|---|---|---|---|---|---|---|
| **Hoa** | -0.08*** | | | | | | | | | |
| Unexplained | 38.5% | | | | | | | | | |
| **Thái** | 0.50*** | 0.57*** | | | | | | | | |
| Unexplained | 52% | 0.87% | | | | | | | | |
| **Tày** | 0.50*** | 0.57*** | 0.003 | | | | | | | |
| Unexplained | 46.5% | 3.2% | 228% | | | | | | | |
| **Mường** | 0.63*** | 0.70*** | 0.13*** | 0.13*** | | | | | | |
| Unexplained | 50% | -1.45% | -2.4% | 2.17% | | | | | | |
| **Khơ-me** | 0.08*** | 0.16*** | -0.42*** | -0.42*** | -0.55*** | | | | | |
| Unexplained | 58.3% | 32.8% | -0.18% | -0.86% | -0.12% | | | | | |
| **Nùng** | 0.64*** | 0.72*** | 0.15*** | 0.15*** | 0.02 | 0.56*** | | | | |
| Unexplained | 62.3% | 32.8% | 3.7% | 30.3% | 60.6% | 1.04% | | | | |
| **Hmong** | 1.54*** | 1.62*** | 1.04*** | 1.04*** | 0.91*** | 1.46*** | 0.90*** | | | |
| Unexplained | 57.9% | 5.4% | 0.9% | 22.4% | 0.97% | -0.07% | 10.23% | | | |
| **Dao** | 1.04*** | 1.12*** | 0.55*** | 0.55*** | 0.42*** | 0.96*** | 0.40*** | -0.50*** | | |
| Unexplained | 72% | 9.8% | -4.0% | 25.7% | 2.36% | 0.61% | 20.5% | 21.1% | | |
| **Ê Đê** | -0.15*** | - 0.07** | -0.64*** | -0.64*** | -0.77*** | -0.23*** | -0.79*** | -1.69*** | -1.19*** | |
| Unexplained | 61.1% | 9.3% | 1.9% | 3.9% | 2.9% | 1.12% | 5% | 7.7% | 0.64% | |
| **Gia Rai** | 0.35*** | 0.42*** | -0.15*** | -0.15*** | -0.28*** | 0.27*** | -0.30*** | -0.50*** | -0.70*** | -0.70*** |
| Unexplained | 117% | 5.09% | 23% | 29% | 15.6% | 1.65% | 24.2% | 21.1% | -3.4% | -3.4% |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

When we look at the decomposition by sector (Table 12), we find that the biggest earning gaps can be attributed to differences between sectors. In the public sector, both gender and ethnic earnings gaps are very small, at 0.02 and -0.07, respectively. It reflects the more structured payment frame in the public sector. In which, the salaries for public servants are calculated by multiplying a fixed base salary regulated by the Government with a co-efficient that varies from person to person, depending on seniority and position. An egalitarian pay structure is applied by the Government to ensure the small differential between the lowest and highest paid. We can also note that the ethnic earnings gap becomes negative, which means that ethnic minorities receive higher earnings than Kinh and only a small part of this gap, 14.6 percent, remains unexplained. The ethnic earning gap is very low in the private formal and informal sectors; however, it is extremely high in agriculture.

Regarding the gender earning gaps, they are relatively similar in the private formal, informal, and agricultural sectors. The finding of gender gap in private sector is consistent with other previous studies on Vietnam labour market (Liu 2004a; Pham and Reilly, 2007). The downsizing of State-owned enterprises affected women more than men since they were overrepresented in public enterprises. The downsizing made them suddenly redundant and more exposed to pay gaps in the less protected private sector (Liu, 2004b).

No other study has explored gender earning gaps in the informal sector of Vietnam so far; however, our finding of large gender earning gaps particularly in this sector is in line with studies

in African countries (Nordman et al., 2011). Since the differences in observable characteristics could not explain much of this gap, other factors might be involved. It is likely that women work in a lower paid segment of the informal sector since the higher paid jobs are more physically demanding. Indeed, manufacturing and construction accounts for the highest proportion of jobs (43 percent) in the informal sector (Cling et al., 2011). Besides, it is possible that women choose to work in lower productivity, but more flexible jobs in order to have more free time to take care of their children and housework. This also explains the relatively large gender earning gap in informal sector: in Vietnam, men in rural areas are better paid than most women because they take on different (heavier and more dangerous) tasks (World Bank, 2012). Another possible explanation could be the physical capital needed to start one's business as an informal street vendor. Women, who have lower physical capital, could be excluded from this high paying informal activity.

**Table 12a: Neumark decomposition of gender and ethnic earnings gaps by sector**

| | **Gender** | | | | **Ethnicity** | | | |
|---|---|---|---|---|---|---|---|---|
| | Gap | Explained | Unexpl. | Unexpl.(%) | Gap | Explained | Unexpl. | Unexpl.(%) |
| Public | 0.02*** | 0.03 | -0.01 | -65.8 | -0.07*** | 0.06 | -0.14 | 184.6 |
| Private | 0.27*** | 0.05 | 0.23 | 83.5 | 0.06*** | -0.02 | 0.08 | 138.9 |
| Informal | 0.26*** | 0.02 | 0.24 | 93.3 | 0.06*** | -0.01 | 0.07 | 116.8 |
| Agriculture | 0.25*** | 0.14 | 0.11 | 44.1 | 0.43*** | 0.32 | 0.11 | 25.3 |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

**Table 12b: Neumark decomposition of gender earnings gaps in the private and informal sectors**

| **Private formal sector** | | | | | **Informal sector** | | | |
|---|---|---|---|---|---|---|---|---|
| | Gap | Explained | Unexpl. | Unexpl.(%) | Gap | Explained | Unexpl. | Unexpl.(%) |
| Wage earners | 0.24*** | 0.04 | 0.22 | 84.1 | 0.45*** | 0.06 | 0.39 | 87.3 |
| Self-employed | 0.31*** | 0.04 | 0.27 | 88.5 | 0.21*** | 0.034 | 0.17 | 83.8 |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

We further investigate earnings gaps among subgroups of wage earners and non-wage earners (including the self-employed and family workers) in the formal and informal sector. The decomposition is presented in Table 12b. It is seen that in the formal sector, the gender gap is slightly narrower for wage workers (0.24) compared to non-wage workers while in the informal sector, the gap for wage worker is twice the size of the gap for non-wage workers. In all cases, the unexplained part accounts for more than 80 percent of the gaps. These results reinforce the idea that in the informal sector, employers have more wage-setting power, leaving room for discrimination. Another explanation could rely on the more competitive product market in the informal sector. In the case of Vietnam, construction accounts for one third of the informal sector and most of the workers are registered as wage workers in this industry (79 percent of workers in construction are wage workers in 2007). Yet, being a wage worker in construction

does not have the same meaning as in other sectors because most of them are in fact at the frontier between self-employment and wage-labour. They are closer from self-employed by hiring their work force at the mercy of private households.

*Full decomposition of gender and ethnic earnings gaps*

Following the raw decompositions in the previous section, we now turn to decompositions within each of the four segments of the Vietnamese labour market (public, formal, informal and agriculture sectors). Those four sectors differ in important characteristics such as job seasonality, wage structures, etc., which could potentially affect the earnings gaps.

Table 13 presents the full decomposition of gender earnings gaps by sectors. Within-sector differences in earnings explain 96.3 percent of the gender earnings gap, of which 26.1 percent is attributed to differences in characteristics between male and female, including differences in educational level, experiences, marital status, and regional differences. The large positive differences in returns to characteristics between men and women imply that men benefit from "nepotism" while women face "discrimination" within sectors. Still, differences in sectoral allocation between men and women only account for 3.6 percent of the gender earnings gaps; most of the differences in earnings are explained by differentials in characteristics. After correcting for the sectoral selection, the gender earnings gap decreases noticeably to 0.119, of which 93 percent is due to within sector differences and nearly half of this difference is explained by differences in characteristics. The remained part is due to differences in returns to endowments between men and women.

**Table 13: Full decomposition of gender earnings gap**

| | Without correcting for selectivity | | Correcting for selectivity | |
|---|---|---|---|---|
| **Gender earnings gap =log(Male group earning) – log(Fem. group earning)** | **0.224*** ** | **%** | **0.119*** ** | **%** |
| Difference due to within sector differences in earnings attributable to | | | | |
| Characteristics | 0.058 | 26.1 | 0.067 | 56.4 |
| Deviation in male returns | 0.078 | 34.8 | 0.023 | 19.5 |
| Deviation in female returns | 0.079 | 35.4 | 0.022 | 18.2 |
| **Sub-total** | **0.215** | **96.3** | **0.112** | **94.1** |
| Difference due to differences between sectoral allocation in earnings attributable to | | | | |
| Characteristics | 0.017 | 7.8 | 0.023 | 19.0 |
| Deviation in effect of characteristics on male location | -0.004 | -1.7 | -0.010 | -8.5 |
| Deviation in effect of characteristics on female location | -0.006 | -2.5 | -0.005 | -4.6 |
| **Sub-total** | **0.007** | **3.6** | **0.07** | **5.9** |

Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

A full decomposition of the ethnic earnings gap is presented in Table 14. In contrast to the gender earnings gap, most of the ethnic earnings gap (about two thirds) is due to differences in sectoral allocation between ethnic majority and minority groups, which can be explained by

differences in characteristics. This finding mirrors our previous findings showing that the ethnic earnings gaps can be explained almost entirely by the minority's high participation in the agricultural sector. More than half of the ethnic earnings gaps could be reduced if ethnic majority and minority groups were allocated equally in the labour market. Within sectors, differences in characteristics explain half of the gap; the other main contribution to the within sector gap is the difference in actual and neutral returns to ethnic minority characteristics. In the correction model, after recalculate the average earning offer within each sector, we find a slightly higher ethnic gap and within sector gap, the deviation in ethnic minority returns to characteristics accounts for 21 percent Thus, it suggests that returns to endowments of ethnic minority are still lower than Kinh counterparts.

**Table 14: Full decomposition of ethnic earnings gap**

| | Without correcting for selectivity | | Correcting for selectivity | |
|---|---|---|---|---|
| **Ethnic earnings gap =log(Maj. group earning) – log(Min. group earning)** | **0.525*** ** | **%** | **0.595*** ** | **%** |
| Difference due to within sector differences in earnings attributable to | | | | |
| Characteristics | 0.088 | 16.7 | 0.136 | 22.9 |
| Deviation in majority group returns | 0.014 | 2.8 | 0.020 | 3.4 |
| Deviation in minority groups returns | 0.064 | 12.3 | 0.125 | 21.1 |
| **Sub-total** | **0.167** | **31.7** | **0.281** | **47.5** |
| Difference due to differences between sectoral allocation in earnings attributable to | | | | |
| Characteristics | 0.271 | 51.5 | 0.255 | 42.9 |
| Deviation in effect of characteristics on majority group location | 0.007 | 1.2 | 0.001 | 0.2 |
| Deviation in effect of characteristics on minority groups' location | 0.081 | 15.6 | 0.056 | 9.4 |
| **Sub-total** | **0.359** | **68.3** | **0.312** | **52.5** |

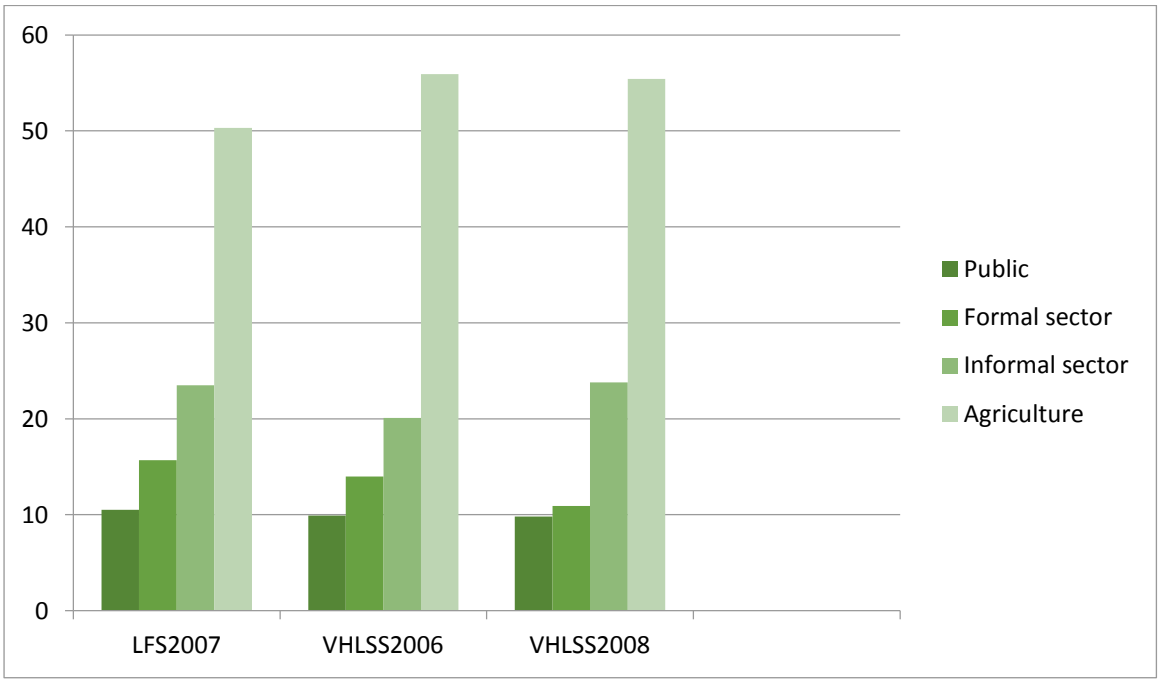
Source: LFS2007, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

## 6 Robustness check

As the results are novel, we conduct robustness checks to demonstrate their validity. As mentioned in section 3, LFS2007 is the most suitable dataset to examine our research questions; however, the drawback is that we only use data for one year. Thus, an alternative dataset is employed and similar analyses are implemented. The most relevant and available dataset for our robustness checks is the Vietnam Household Living Standard Survey (VHLSS). VHLSS is a large-scale and high quality survey of the Vietnamese population implemented every two years. Since we use the LFS in 2007, we will employ the VHLSS in 2006 and 2008 as alternative samples. Questionnaires about income and employment in VHLSS are limited to random subsamples of 39,071 individuals in 2006 and 38,253 individuals in 2008.

Regarding sector classification, compared to LFS2007, VHLSS is limited in capturing the concept of the informal sector. In the VHLSS, information about whether unincorporated enterprises register/have a business license is only available for employers/self-employed workers. For wage workers, informal employment is captured by information about the social security contributions of their job. Thus, we follow the informality proxy construction by Nguyen et al., (2013), according to which workers belong to the informal sector if they are self-employed in an unregistered business or if they are wage-workers without social security. Even though the definition of the informal sector is not consistent between the LFS and the VHLSS, we find similar sectoral compositions of the labour market (Figure 1). We therefore think that the VHLSS can be used confidently for robustness checks.

**Figure 1. Structure of the labor market (proportion of each sector in the labor market)**




Source: LFS2007, VHLSS2006, VHLSS2008, GSO; authors' calculations.

Neumark decompositions method (Table 15) provide consistent results for the gender earning in both 2006 and 2008, of which almost two thirds are left unexplained. Ethnic earning gaps are even larger when estimated using the VHLSS but only one fourth of this gap is explained by the control variables. We also decompose the earning gaps by sector (Table 16) and apply the full decomposition to test within and between sector gaps (Tables 17 & 18). Similar to our results using the LFS, the gender earning gap is low in the public sector but similar among the remaining sectors; the ethnic earnings gap is highest in agriculture. Full decompositions again confirm that differences in sectoral allocation affect the earning gaps: within-sector differences are the main drivers of the gender earning gaps, but less significant in explaining the ethnic gap.

**Table 15: Neumark decomposition of gender and ethnic earnings gaps using VHLSS**

| | | Raw earning gap | Explained | Unexplained | % Unexplained |
|---|---|---|---|---|---|
| Gender earnings gaps | 2006 | 0.177*** | 0.057 | 0.120 | 67.70 |
| | 2008 | 0.162*** | 0.058 | 0.104 | 64.00 |
| Ethnic earnings gaps | 2006 | 0.633*** | 0.476 | 0.157 | 24.83 |
| | 2008 | 0.611*** | 0.454 | 0.157 | 25.8 |

Source: VHLSS2006, VHLSS2008, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

**Table 16: Neumark decomposition of gender and ethnic earnings gaps by sector using VHLSS**

| | | Gender Gap | Explain | Unexpl. | Unexpl. (%) | Ethnic Gap | Explain | Unexpl. | Unexpl. (%) |
|---|---|---|---|---|---|---|---|---|---|
| Public | 2006 | 0.056* | 0.184 | 0.037 | 66.8 | 0.153*** | 0.178 | -0.0249 | -16.3 |
| | 2008 | -0.014 | 0.033 | -0.046 | 343 | 0.215*** | 0.166 | 0.048 | 22.6 |
| Private | 2006 | 0.211*** | 0.041 | 0.169 | 80.3 | 0.172*** | 0.129 | 0.042 | 24.49 |
| | 2008 | 0.290*** | 0.078 | 0.212 | 73.05 | 0.342*** | 0.266 | 0.076 | 22.4 |
| Informal | 2006 | 0.266*** | 0.041 | 0.225 | 84.49 | 0.142*** | 0.079 | 0.623 | 43.9 |
| | 2008 | 0.256*** | 0.034 | 0.222 | 86.7 | 0.180*** | 0.082 | 0.098 | 54.2 |
| Agriculture | 2006 | 0.121*** | 0.082 | 0.039 | 32.6 | 0.638*** | 0.493 | 0.145 | 22.66 |
| | 2008 | 0.081*** | 0.752 | 0.005 | 7.07 | 0.594*** | 0.461 | 0.132 | 22.3 |

Source: VHLSS2006, VHLSS2008, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

**Table 17: Full decomposition of gender earnings gap using VHLSS**

| **Gender earnings gap =log (Male group earning) – log (Fem. group earning)** | **2006** | **0.177*** ** | **%** |
|---|---|---|---|
| | **2008** | **0.162*** ** | |
| Difference due to within sector differences in earnings attributable to | | | |
| Characteristics | 2006 | 0.059 | 33.1 |
| | 2008 | 0.048 | 29.4 |
| Deviation in male returns | 2006 | 0.048 | 27.0 |
| | 2008 | 0.044 | 27.2 |
| Deviation in female returns | 2006 | 0.050 | 28.3 |
| | 2008 | 0.047 | 29.3 |
| **Sub-total** | **2006** | **0.157** | **88.4** |
| | **2008** | **0.139** | **85.9** |
| Difference due to differences between sectoral allocation in earnings attributable to | | | |
| Characteristics | 2006 | 0.016 | 9.1 |
| | 2008 | 0.016 | 9.6 |
| Deviation in effect of characteristics on male location | 2006 | 0.003 | 1.5 |
| | 2008 | 0.005 | 2.8 |
| Deviation in effect of characteristics on female location | 2006 | 0.002 | 0.9 |
| | 2008 | 0.003 | 1.7 |
| **Sub-total** | **2006** | **0.021** | **11.5** |
| | **2008** | **0.024** | **14.1** |

Source: VHLSS2006, VHLSS2008, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

**Table 18: Full decomposition of ethnic earnings gap using VHLSS**

| **Ethnic earnings gap =log(Maj. group earning) – log(Min. group earning)** | **2006** | **0.640*** ** | **%** |
|---|---|---|---|
| | **2008** | **0.617*** ** | |
| Difference due to within sector differences in earnings attributable to | | | |
| Characteristics | 2006 | 0.292 | 45.6 |
| | 2008 | 0.281 | 45.6 |
| Deviation in majority group returns | 2006 | 0.032 | 5.0 |
| | 2008 | 0.032 | 5.2 |
| Deviation in minority group returns | 2006 | 0.098 | 15.4 |
| | 2008 | 0.114 | 18.4 |
| **Sub-total** | **2006** | **0.422** | **66.0** |
| | **2008** | **0.427** | **69.2** |
| Difference due to differences between sectoral allocation in earnings attributable to | | | |
| Characteristics | 2006 | 0.164 | 25.6 |
| | 2008 | 0.146 | 23.6 |
| Deviation in effect of characteristics on majority group location | 2006 | 0.002 | 0.3 |
| | 2008 | 0.001 | 0.2 |
| Deviation in effect of characteristics on minority group location | 2006 | 0.050 | 8.0 |
| | 2008 | 0.041 | 6.7 |
| **Sub-total** | **2006** | **0.216** | **34** |
| | **2008** | **0.188** | **30.8** |

Source: VHLSS2006, VHLSS2008, GSO; authors' calculations.
Note: ***: significant at 1%; **: significant at 5%;*: significant at 10%.

## 7 Conclusion

This paper uses the 2007 Labour Force Survey to estimate and analyze the gender and ethnic earnings gaps in Vietnam. Previous research had brought up the potential for discrimination against women and ethnic minorities. Their studies focused mostly on wage employment and disregarded the informal sector. However, in Vietnam's labour market, women's participation in the informal sector is substantial and a majority of ethnic minority workers are engaged in non-waged occupations. Previous studies were thus severely limited in analyzing the gender and ethnic gaps and did further fail to disaggregate the different ethnic groups. We overcome these aforementioned issues by including the self-employed and agricultural workers in estimating the gender and ethnic earnings gaps and investigate further the heterogeneity in the gaps among different ethnic groups.

We start from Mincerian earnings equation, using a multinomial logit model to correct for selection into different labour market sectors. Without correcting for selection bias, ethnic minority people earn 22 percent less than Kinh, the majority group, while the gender gap is 16.4 percent We find a great reduction in the ethnic earnings gap, which reduces 2.6 percent when looking at non-farm subsample only. Correcting for selection with our multinomial logit model shows that the earning gap between Kinh and ethnic minority group is 48 percentage points in the agricultural sector while it is 18 percentage points when they are engaged in non-agricultural formal and informal employment. Our results for agriculture and nonfarm occupations further suggest that the remaining gap in farming and rural activities is probably more due to subtle differences in access to markets and land quality than due to discrimination, whether employers or self-driven. However, we cannot eliminate the effect of past discrimination and accumulative discrimination on the current disparities in endowments between ethnic minority and majority groups. When ethnic minority was discriminated against and pushed into the remote and difficult area in the past, they would be trapped in areas with poor endowments and the disparities would even increase over time. Moreover, female ethnic minority members have lower earnings than Kinh or non-Kinh men in all but the public and agriculture sectors. In all models, the earning gap coefficients are extremely heterogeneous. Hmong, the poorest and least assimilated ethnic minority in Vietnam, face the largest earnings gap. The heterogeneity in ethnic earnings gaps raises the need of targeted public policies rather than a "one size fits all" policy for all ethnic minority groups.

Decomposition results reflect the different pattern observed between gender and ethnic earnings gaps. Using a multinomial logit model to correct for selectivity into different employment sectors, we find that within-sector differences explain most of the gender earning gaps and between-sector differences favour male workers in terms of characteristics. While the gender earnings gap is small, the high proportion of the unexplained gap especially in the informal sector is noteworthy, where the gap is larger for wage earners with 87 percent of the gap left unexplained. Thus, our finding supports the common view that the gender earnings gaps are likely to be the result of employer-based discrimination. The ethnic earnings gap, however, is explained almost equally by within- and between-sector differences. This corresponds to our finding that the ethnic earnings gaps are captured almost entirely by the predominant participation of ethnic minorities in the agriculture sector. Besides, in contrast to our gender gap

results, we find that differences in characteristics make up about two third of the earnings gap between Kinh and non-Kinh. These findings confirm the existence of ethnic earnings gap in Vietnam (Pham and Reilly, 2009) and the necessity to strengthen policies to increase educational level and opportunities to access non-farm jobs for ethnic minority groups. Our findings are, to some extent, similar to what has been found in Latin American countries, namely that the ethnic earnings gap is higher than the gender earnings gap and that occupational segregation plays an important role (Atal et al., 2009), while it is not the case in some African countries (Nordman et al., 2011).

**References**


Appleton, S., Hoddinott, J., and Krishnan, P. 1999. 'The Gender Wage Gap in Three African Countries'. *Economic Development and Cultural Change*, 47(2): 289-312.

Arrow, K.J. 1973. 'The Theory of Discrimination'. In Ashenfelter, O., Rees, A. (eds), *Discrimination in Labor Markets*. Princeton University Press.

Asian Development Bank. 2004. 'The Factor Markets in Vietnam'. Discussion Paper 5, Making Markets Work Better for the Poor. Hanoi: Asia Development Bank.

Atal, J., Nopo, H., and Winder, N. 2009. 'New Century, Old Disparities: Gender and Ethnic Wage Gaps in Latin America'. Working Paper 109. Inter-American Development Bank.

Baulch, B. 2010. 'Ethnic Minority Poverty in Vietnam'. Working Paper 169. Chronic Poverty Research Centre.

Baulch, B., Chuyen, T., Haughton, D., and Haughton, J. 2007. 'Ethnic minority development in Vietnam'. *The Journal of Development Studies*, 43(7): 1151-76.

Baulch, B., Pham, H., & Reilly, B. 2012. 'Decomposing the Ethnic Gap in Rural Vietnam, 1993–2004'. *Oxford Development Studies*, 40/1: 87-117.

Baulch, B. and Vu, D. H. 2011. 'Poverty dynamics in Vietnam, 2002 – 2006'. In Baulch, B. (ed.), *Why poverty persists: poverty dynamics in Asia and Africa*. Gloucester: Edward Elgar.

Becker, G.S. 1957. 'The Economics of Discrimination'. *University of Chicago Press*.

Blau, F., and Kahn, L. 2003. 'Understanding International Differences in the Gender Pay Gap'. *Journal of Labor Economics*, 21(1): 106-44.

Bourguignon, F., Fournier, M., and Gurgand, M. 2007. 'Selection Bias Correction based on the Multinomial Logit Model: Monte Carlo Comparisons'. *Journal of Economic Surveys*, 21(1): 174-205.

Canelas, C., and Salazar, S. 2014. 'Gender and ethnic inequalities in LAC countries'. *IZA Journal of Labor & Development*, 3(1): 18.

Cling J. P., Marouani M. A., Razafindrakoto M., Robilliard A.S., and Roubaud F. 2009. 'The Distributive impact of Vietnam's accession to the WTO'. *International Economy* (forthcoming).

Friedman, J., Goodkind, D., Cuong, B., and Anh, T. 2001. 'Work and Retirement among the Elderly in Vietnam'. *Research on Aging*, 23(2): 209-32.

Heckman, J. 1979. 'Sample Selection Bias as a Specification Error'. *Econometrica*, 47(1): 153.

Gallup, J. 2004. 'The Wage Labor Market and Inequality in Vietnam'. In Glewwe, P., Agrawal, N. and Dollar, D. (ed.)*, Economic Growth, Poverty, and Household Welfare in Vietnam*. Washington DC: World Bank.

GSO. 2008. 'Report on Labour Force and Employment Survey'. Hanoi: Statistical Publishing House.

Heckman, J. 1979. 'Sample Selection Bias as a Specification Error'. *Econometrica,* 47: 153-61.

Hoang, H., Pham, G., Tran, M., and Hansen, H. 2007. 'Ethnicity and poverty reduction'. In Nguyen, T. and Hansen, H. (eds) *Markets, Policy and Poverty Reduction in Vietnam*. Hanoi: Vietnam Academy of Social Sciences/National Political Publishing House.

Lee, L. F. 1983. 'Generalized econometric models with selectivity'. *Econometrica,* 51: 507-12.

Liu, A. 2004a. 'Sectoral gender wage gap in Vietnam'. *Oxford Development Studies*, 32(2): 225-39.

Liu, A. 2004b. 'Gender wage gap in Vietnam: 1993 to 1998'. *Journal of Comparative Economics*, 32(3): 586-96.

McElwee P. 2004. 'Becoming Socialist or Becoming Kinh? Government Policies for Ethnic Minorities in the Socialist Republic of Vietnam'. In Duncan C., *Civilizing the Margins*. Ithaca: Cornell University Press.

Mincer J. 1974. 'The Human Capital Earnings Function'. In *Schooling, Experience and Earnings*. National Bureau of Economic Research.

Neumark, D. 1988. 'Employers' Discriminatory Behavior and the Estimation of Wage Discrimination'. *The Journal of Human Resources*, 23(3): 279-95.

Nguyen, H., Nordman, C., and Roubaud, F. 2013. 'Who Suffers the Penalty?: A Panel Data Analysis of Earnings Gaps in Vietnam'. *Journal of Development Studies*, 49(12): 1694-710.

Ñopo, H., Daza, N., Ramos, J. 2011. 'Gender Earnings Gaps in the World Gender Earnings Gaps in the World'. IZA Discussion Papers 5736.

Nordman, C., Robilliard, A., and Roubaud, F. 2011. 'Gender and ethnic earnings gaps in seven West African cities'. *Labour Economics*, 18: 132-45.

Pham, H., and Reilly, B. 2009. 'Ethnic wage inequality in Vietnam'. *International Journal of Manpower*, 30(3): 192-19.

Pham, T., and Reilly, B. 2007. 'The gender pay gap in Vietnam, 1993–2002: A quantile regression approach'. *Journal of Asian Economics*, 18/5: 775-808.

Phelps, E. 1972. 'The statistical theory of racism and sexism'. *American Economics Review*, 62(4): 659–61.

Phung, D.T., and Do, T.T. 2014. '54 ethnic groups: Why different?'. Hanoi: Mekong Development Research Institute.

Razafindrakoto, M., Roubaud, F. 2007. 'Towards a Better Monitoring of the Labor Market'. In *Vietnam Development Report 2008: Social Protection, Joint Donor Report*. Hanoi: World Bank.

Razafindrakoto, M., Roubaud, F., Nguyen, H. C. 2011. 'Vietnam Labor Market: An Informal Sector Perspective'. In Nguyen, D. T. (ed.), *Vietnam Annual Economic Report 2011: The Economy at a Crossroad*. Hanoi: Tri Thuc Edition.

Roubaud, F. 1994. 'La Question Ethnique sur le Marché du Travail à Youndé: Discrimination ou Solidarité?'. DIAL Working Paper DT/1994-13. Paris : DIAL.

Taylor, P. 2004. 'Introduction: social inequality in a socialist state'. In Taylor, P. (ed.) *Social inequality in Vietnam and the challenges to reform*. Singapore: Institute of Southeast Asia Studies.

Tzannatos, Z. 1999. 'Women and Labor Market Changes in the Global Economy: Growth Helps, Inequalities Hurt and Public Policy Matters'. *World Development*, 27(3): 551-69.

Van de Walle, D., and Gunewardena, D. 2001. 'Sources of ethnic inequality in Vietnam'. *Journal of Development Economics*, 65(1): 177-207.

Weichselbaumer, D., Winterebmer, R. and Zweimuller, M. 2008. 'Market Orientation and Gender Wage Gaps: an International Study'. *Kyklos, Wiley Blackwell,* 61: 615–35.

World Bank. 2012. 'Vietnam poverty assessment: Well Begun, Not Yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges'. Washington DC: World Bank.

World Bank. 2011. 'Vietnam Country Gender Assessment'. Washington DC: World Bank.

# Chapter 3

# Ethnicity and risk sharing network formation

## Evidence from rural Vietnam

Co-authored with Laure Pasquier-Doumer, and Camille Saint-Macary

**Abstract**
Ethnic gaps remain a persistent challenge for Viet Nam. This paper aims at exploring the mechanisms driving to ethnic inequality by analysing the formation of risk sharing networks in rural Viet Nam. After first showing important differences in terms of size and similarity attributes between the networks of ethnic minority and of Kinh households, we then explore homophily in link formation to explain such differences. We disentangle baseline homophily which derives from the different local distribution of ethnic groups, from inbreeding homophily which results from preferences to form links with similar people, and is associated to cultural and social distance between ethnic groups. We thus estimate various models to investigate the determinants of risk sharing networks attributes, among which ethnicity, geography, individual and household characteristics, as well as ethnic composition. Results show that ethnic minorities have smaller and less diversified networks than the majority. Ethnicity plays a direct role on network formation through the combination of baseline and inbreeding homophily mechanisms. Inbreeding homophily is found to be stronger among the Kinh majority, leading to the exclusion of ethnic minorities from Kinh networks, which are supposed to be more efficient to cope with covariant risk because they are more diversified in the occupation and location of their members. Our evidence suggests that inequalities among ethnic groups in Viet Nam are partly rooted in the cultural and social distances between them.

*Keywords: Risk-sharing network, homophily, ethnic gap, Vietnam*

*JEL classification: O12, I31, D85*

**Acknowledgements:** The authors would like to thank the participants in the VARHS project workshop in Helsinki (February 2018), Vietnam Economist Annual Meeting in Hanoi (June 2018), the 6th Atelier DIAL-GREThA développement in Paris (June 2018), the Sustainability and Development Conference in Michigan (November, 2018), and the GDRI 2018 Conference in Clermont-Ferrand (November 2018) for the insightful comments and questions provided.

## 1 Introduction

Accounting for only 15 per cent of the population but 70 per cent of Vietnam's extreme poor (World Bank, 2012), ethnic minorities lag behind the Kinh majority despite Government's effort to narrow the ethnic gap through numerous poverty alleviation programmes. Poverty of ethnic minorities remains a persistent challenge for the country.

The reasons for the existence and the persistence of this gap—despite the fast economic growth and high pace of poverty reduction experienced in Vietnam over the last decade—have been explored in the literature. Research highlights the existence of inequalities in the endowments between ethnic groups with a strong emphasis on geographical factor, but also shows differences in the returns to these endowments between Kinh and non Kinh, mainly regarding education, land holding, and access to credit (Baulch et al. 2007, 2012; Imai et al., 2015; Van de Walle and Gunewardena, 2001). In a recent example, Singhal and Beck (2015) show that the welfare gap between Kinh and non Kinh is explained by disparities in agricultural land quality, in the ownership of certificates, by lower diversification of non-farm jobs, and limited credit access, etc., even when the location factor is controlled for. In addition, this study also provides some descriptive evidence about ethnic segmentation in social networks.

In this paper, we investigate the question of the ethnic inequality using the angle of risk sharing and social networks formation. We are motivated by the fact that risk, and the necessity to cope with it, is a pervasive feature of rural economic life, particularly in Vietnam's mountainous areas, where ethnic minorities mostly live. There, the prevalence of risk is accentuated by climate change, by difficulties to cultivate fragile uplands, and by high transport costs, and is combined with low access to basic infrastructure, public services, and formal insurance (Van de Walle and Gunewardena, 2001; Imai et al., 2011; Tran, 2015). In this context of missing or failing market, the literature demonstrates the importance of the prevalence of bad shocks to explain the persistence of poverty (Morduch, 1994), and highlights risk sharing networks as a dominant means to cope with shocks for rural poor (Rockenbauch and Sakdapolrak, 2017).

Ethnic minorities in Vietam mostly live in mountainous areas while the Kinh majority is present all over the country, but more concentrated in the lowland areas. This spatial segregation leads to important differences in endowments between ethnic groups, as well as in occupations, and in risk sharing networks. Yet, we expect that in addition to these important factors, ethnicity also plays a direct role in the formation of social networks (Grimard, 1997) that leads to further dissimilarities that may affect the ability of ethnic minority and majority groups to cope with shocks.

Literature on homophily in link formation suggests that having similar characteristics like ethnicity is a strong predictor of two individuals being connected. Assortativity patterns could be a main mechanism of ethnic segregation (Jackson et al., 2016; Currarini et al., 2009; McPherson et al., 2001). Segregation can occur because of a lack of opportunities and/or low preferences to interact with people from different ethnic group. The opportunities for a group to interact with another group, defined as baseline homophily, depends on the size of the ethnic groups. Because the Kinh majority group constitutes 85% of the population while the 53 other ethnic groups account for the remaining part, opportunities to exchange with people from different ethnic group may differ a great deal across majority and minority ethnic groups. Opportunities to form a link also depend on the local distribution of ethnic groups. Preference to interact with people from the same ethnic group is called inbreeding homophily in the literature. Prejudice and stereotypes regarding ethnic minority people are prevalent in Vietnam, suggesting the existence of social and cultural distance across ethnic groups and differentiated preferences to interethnic interactions. It could lead to higher exclusion of non-Kinh people from Kinh risk sharing

network than predicted by their demography and local distribution. Thus, local distribution of ethnic groups and social distance across ethnic groups may explain differences in similarity patterns across ethnic group social networks.

In this paper, we first describe various attributes of the risk sharing networks of households from the majority ethnic group and from minority ethnic groups and their differences. Ethnic minority groups are embedded in smaller risk sharing networks, which may imply a lower capacity to pool the risks. In addition, their network is characterized by higher similarity patterns in terms of occupation and location. Being connected with people with the same economic activity and living in the same village is expected to be less efficient when coping with covariant shocks. Indeed, covariant shocks, by definition, similarly affect people having the same activity and residing in the same geographical area. Starting from these observations, this paper aims at understanding which mechanisms are driving to theses observed differences, other than differences in geography and in socio-economic conditions that we control through fixed effects and other control variables. Here, we focus on homophily as a main mechanism of differentiation, and disentangle baseline homophily that is explained by differences in the local distribution of ethnic groups, from inbreeding homophily that results from preferences, and from cultural and social distance between ethnic groups. Providing empirical evidence on the impact of ethnicity on network formation is challenging since it requires identifying exogenous variations in ethnic composition at the local level. Our approach is to provide an in-depth descriptive assessment on the formation of risk sharing networks by characterizing the differences in network's structure between the ethnic minority and majority groups, and by analysing multivariate correlations between ethnic composition within communes, geography, and risk sharing network features.

The contribution of this paper is threefold. First, it provides evidence in a developing country on the role played by ethnicity in the formation of risk sharing networks, and distinguishes various mechanisms through which it acts. The literature on this issue is up to now scarce. Yet, better understanding the differences in risk sharing networks between ethnic groups is likely to provide valuable information to address the ethnic gap which is a challenging issue in the context of Vietnam. Second, it provides a representative picture of the risk sharing formation process on a large scale (at least provincial level), going beyond most of studies on risk sharing network, that limit the validity of their results to some unrepresentative villages in a country. In addition, link formation outside the village is taken into consideration. This is all the more important since several studies have shown that the village is not the relevant risk sharing unit (Grimard, 1997, Jaramillo et al., 2015). Finally, by combining census and spatial data, along with household survey data, it links information on ethnicity composition and geographical environment at a very detailed level that is the commune. To the best of our knowledge, this is the first time it has been done for Vietnam.[12]

The rest of our paper is organized as follows: section 2 presents the conceptual background of the paper; the dataset and definitions are shown in section 3. Section 4 then provides our descriptive analysis. We present our estimation strategy and results in section 5 and briefly conclude in section 6.

[12] In Nguyen et al. (2017) for example, ethnic composition at the commune level is obtained from VHLSS household survey sample, which is not representative at such geographical level. In addition, the measurement of ethnic composition is oversimplified through a dummy variable taking the value 1 if at least two people have different ethnic group in the commune. The commune is thus defined as mixed commune but nothing is known about the distribution of ethnic groups in the commune.

## 2 Conceptual background

### 2.1 Importance of risk sharing network in the strategies to cope with risk

It is widely known that risks are highly prevalent in rural areas of developing countries and risk is considered a cause of poverty and its persistence. Indeed, people in rural and remote areas face high incidence of diseases, environmental hazards, and business risk due to self-employment activities (Fafchamps et al., 1998, Fafchamps and Lund, 2003; De Weerdt and Dercon, 2006). As a consequence, shocks, and the risks that they occur, are central to people's lives in rural areas. On the other hand, they do not have sufficient wealth to absorb the negative shocks while the formal mechanisms and institutions to deal with risks are limited in those areas.

The adverse impacts of high exposure to shocks, however, could be compensated or mitigated by households' strategy to reduce their impact. Literature has well documented the set of these strategies (Dercon, 2002), particularly risk sharing arrangements, where households share the risk with others through transfers in gift, money and/or labour.

Risk sharing has been investigated theoretically and empirically in a vast literature, suggesting various forms of risk sharing networks. The theoretical works are based on a Pareto optimal model of risk pooling, conditional on the enforceability of risk sharing agreements/contracts (Coate and Ravallion, 1993; Townsend, 1994; Bramoullé and Kranton, 2007). Different motivations lead households to engage in risk sharing arrangements with other households. While risk-aversion is the main driver of risk sharing network formation, they can be sustained only if informal mechanisms of enforcement are at play, in contexts where reliance on the court system is not possible. Repeated game theory is one kind of possible enforcement mechanism in informal arrangements (Fafchamps, 2011). In the absence of altruism, long-term strategic interactions with expectation for mutual help and reciprocity may be sufficient to guarantee sustainable risk sharing arrangements. It is less true when shocks are persistent. Emotions such as altruism, guilt or shame which are usually embedded in redistributive social norms constitute another enforcement mechanism.

### 2.2 Homophily as central pattern of risk sharing network formation

Another strand of the literature on risk sharing networks investigates the formation of links. In this literature, to which this paper aims at contributing, the question is not why people form links with others, but with whom they choose to connect. One of the most pervasive patterns of link formation is homophily: having similar characteristics is generally a strong and significant predictor of two individuals being connected (McPherson et al., 2001). Although the benefit from sharing income risk is the largest when people have uncorrelated or negatively correlated incomes, and therefore when they have different characteristics, it is widely observed that links are more likely to be formed when social and geographical distance is the lowest. Indeed, distance also raises the cost of establishing and maintaining interpersonal links, and weakens enforcement mechanisms. Numerous empirical studies illustrate this pattern, like Fafchamps and Gubert (2007), who find that informal arrangements in rural Philippines are based on social and geographical proximity.

Homophily may impact risk sharing efficiency through three channels. First, it could make the size of the network smaller than the optimal size as individuals connect mostly with close or similar individuals. Based on the hypothesis of single-person deviation alone, theoretical models predict that larger groups allow for more diversification of shocks, leading to higher gains from

sharing risk (Genicot and Ray, 2003).[13] Second, social and geographical proximity may impede the risk pooling group to diversify their risk, especially when covariant risks are predominant. Neighbours are more likely to face simultaneously a covariant shock like flood or pest making risk sharing arrangements unsustainable. Households sharing the same activity in the same area will be similarly impacted by crop or input price changes. Due to a limited resource endowment, risk sharing networks of a poor ethnic minority community have less capacity to smoothen income variation within this community. Studying income diversification in rural Kerala, Johny et al. (2017) show that caste based homophily negatively affects the risk management strategy: networks with strong intra-caste connections associate with less diversification of income sources. Third, homophily reversely helps to cope with idiosyncratic shocks by lowering transaction costs or information asymmetries, and by strengthening mutual enforcement. Using randomized experiments across 34 Indian villages, Chandrasekhar et al. (2014) demonstrate that socially close individuals maintain high levels of cooperation even when contract enforcement is removed, while more distant individuals do not. Attanasio *et al.* (2012) observe from an experiment conducted in 70 Colombian communities that close individuals are more likely to join the same risk pooling group and they group assortatively on risk attitudes, while socially distant participants group less and rarely assort. They conclude that social homophily is a necessary condition to benefit from advantages to grouping assortatively on risk attitudes.

### 2.3 Baseline and inbreeding homophily to explain ethnic segregation in the networks

Yet, the role of ethnic-based homophily in risk sharing network formation has been poorly documented, especially in developing countries. Existing studies on ethnic-based homophily mostly deal with friendship network formation in developed countries. Notably, Currarini et al. (2009) examine friendship patterns in US high schools. Their theoretical model and empirical estimation show that larger ethnic groups form significantly more friendships per capita. In addition, larger ethnic groups tend to form more ties with people from the same ethnic group and fewer from other ethnic groups than small ethnic groups. McPherson et al. (2001) or Jackson et al. (2016) review other empirical studies demonstrating similar results.

Sociologists have decomposed ethnic based homophily into two mechanisms: the first one, called baseline homophily, relies on ethnic composition and demographic of ethnic groups; the second one, called inbreeding homophily, on preferences. According to baseline homophily, link formation across ethnic groups depends on their relative size. The demography conditions the potential tie pool, i.e the opportunities of a group to interact with another group. The more numerous an ethnic group, the more likely is one of its member to meet and thus to interact with someone from the same ethnic group. As for inbreeding homophily, it refers to a selection process within the set of opportunities to interact, or in other words the preference of interacting with someone from the same ethnic group. As a result, the propensity for ethnic groups to form links with co-ethnic people exceeds their relative fractions in the population. Various reasons for inbreeding homophily can be found in the literature: shared cultural traits and other co-ethnic related factors like language would help people to interact more effectively and on a longer term, to be better able to judge their co-ethnic's characteristics, to use information or rules to maintain the effective cooperation (Alesina and La Ferrara, 2005) or to apply social sanctioning (Khwaja,

[13] However, with coalitional deviations hypothesis, the relationship between risk sharing network size and efficiency becomes unclear, as larger groups can be destabilised by smaller sub-groups that can renege on the informal arrangement, and thant are large enough to provide significant levels of risk sharing (Genicot and Ray, 2003). The empirical literature testing the relationship between size and efficiency is scarce. To our knowledge, only Fitzsimons et al. (2018) analyses this relationship in the context of family risk-sharing networks in Malawi. They find that households where the wife has many brothers achieve worse risk-sharing relative to households where she has fewer brothers.

2009; Miguel and Gugerty, 2005). 'Other-regarding preferences' may be another reason as far as people tend to care more about the welfare of their co-ethnic people than the one of non co-ethnic people (Habyarimana et al., 2007).

### 2.4 Contribution of the paper to risk sharing network literature

From literature on homophily in link formation, mostly relating to developed countries, one might expect that ethnic groups have different risk sharing network structures in terms of size and similarity, which in turn could affect the efficiency of these networks to cope with covariant and idiosyncratic risks. To the best of our knowledge, no study investigates the difference among ethnic groups in terms of the risk sharing networks in developing countries, although this may explain why some ethnic groups are more vulnerable to shocks than others. One exception is Arcand and Jaimovich (2014). Relying on the census of almost complete networks in 59 villages in The Gambia, they investigate whether ethnic diversity within villages affects the structure of networks where land, labour, inputs, and credit are exchanged. In particular, they test if ethnic fragmentation and polarization explain features of the exchange network, which are its density, its compactness, and its clustering, or affect the probability of forming a link with individuals from the same ethnic group. The authors find no evidence of any effects of ethnic diversity on the network architecture or on the probability to form a link. In addition, they find that households belonging to ethnic minorities do not have less economic interactions on average. However, as acknowledged by the authors, this study has been conducted in a very specific context where religion and culture are common to most villagers despite ethnic heterogeneity. This leads the authors to call for other evidence about the effect of ethnic diversity on social networks in context where social distance between ethnic groups is observed. In addition, exchange networks may substantively differ from risk sharing networks as they include effective exchanges and not the potential ones.

This paper aims at contributing to the literature on the risk sharing formation by investigating whether and why ethnic minority groups in Vietnam differ in the structure of their network compared to the ethnic majority group. Structure is approached here by the size and the similarity patterns of the network. The idea is to scrutinize two mechanisms that are baseline homophily and inbreeding homophily. In other words, we try to disentangle the effects of demography of ethnic groups and their local distribution from the effect of social distance between ethnic groups. While doing so, we take into account the heterogeneity of geographical environment explaining differences in the cost of forming and maintaining links.

### 2.5 Relevance and evidence in the context of Vietnam

This question is particularly relevant in the context of Vietnam, as it provides new insights on the persistence of the ethnic gap. Although ethnic gap remains one important challenge of the political agenda, inequalities in risk sharing networks across ethnic groups have not yet been examined, even less their underlying mechanisms. However, scattered and circumstantial evidence suggests that ethnic-based difference in risk sharing networks is an issue in Vietnam. Zylberberg (2010) finds that households from ethnic minorities participate significantly less in risk-pooling in the aftermath of a typhoon than Kinh do. Tran (2015) finds that the Kinh have larger risk sharing networks than households from ethnic minority groups.

In addition, recent migration patterns make the local distribution of ethnic groups of great interest as it presents a large spectrum of situation. Oversimplifying, Kinh people mainly occupied the delta and the coastal areas, while ethnic minorities lived in the highland areas, especially in Northern and Central Vietnam. Until the Independence, mixed communes with

Kinh and ethnic minority people were rare (Déry, 2003).[14] From the Independence but especially after the reunification in 1975, voluntarist resettlement policy has been conducted by the government. Numerous Kinh households have moved to the mountainous areas and lived with the ethnic minorities for the purpose, among others, of integrating ethnic minorities in the country's political path and building national unity (Mc Elwee, 2004; Saint-Macary, 2014). According to the Marxist theory of nationalities, the minorities must, in contact with Kinh majority, change their habits and adopt new practices (Hardy and Nguyen, 2004). Policies of sedentarization of ethnic minority communities practicing slash-and-burn agriculture started to be implemented in the same period (Nguyen *et al.*, 2017). In addition, existence of settlements due to government programmes gave rise to networks and spontaneous migration (Hardy, 2003). It is estimated that between 1976 and the late 90s, up to six million people resettled in the whole country (UNDP, 1998). As a result, the ethnic composition of mountainous areas has completely changed, and communes mixing Kinh with ethnic minority people are a relatively recent phenomenon. Compared to ethnic minority groups, Kinh people benefit from demographic preponderance but also from being spread in the whole territory. As such, their potential tie pool is much larger and more widely spread. According to baseline homophily, one should expect larger risk sharing networks with more co-ethnics and less geographical and occupational similarity.

Finally, sociologists and anthropologists have demonstrated the existence of social distance between Kinh and ethnic minority people in various contexts of Vietnam. The sociologist Vu Hong Phong (2013) conducted field work on interethnic relationship in Muong Khen, a rural town in northwest mountainous Vietnam where Muong, native to the area, and Kinh, internal migrants, are living together. It shows complex interethnic relations resulting from inflow of Kinh migrants from lowland provinces, the settlement of Kinh families in Muong villages, the establishment of multiethnic rural cooperatives, the Vietnam War, and the economic reform (*Đổi mới)* which dismantled the rural cooperatives and urged many Kinh families to leave Muong villages. Although cooperation between Kinh and Muong people can be found, a lot of tensions remain due to competition over resources and stereotypes. According to the author, many Kinh still ascribe 'drinking heavily', 'nepotism' (*cục bộ*) and 'backwardness' (*lạc hậu*) to the Muong; meanwhile, many Muong still think that the Kinh 'look down' at them, and attribute 'lacking in sentiments' and 'cunning' to the Kinh. Some interviews of Kinh people illustrate this social distance: '*When the Kinh came here, they always married Kinh and never married Muong. They considered the Muong mọi [savage], therefore, they did not marry them! I knew they did not marry them because they were proud of themselves. The Muong did not think like that. Only the Kinh thought like that…' (Kinh man, 80 years old, Case 9)"* (Vu Hong Phong, 2013, p.179), or '*though some people in the Muong village are just like us, many are keeping up with their customs. The thing is that I do not want to follow their customs. They are complicated! They are not as simple as ours are. I was tired of following their customs. If I had sworn brotherhood with a Kinh, it would have been easier.' (Kinh man, 36 years old, Case 34)*', (Vu Hong Phong, 2013, p.175). Looking at the prevalence of interethnic relationships which is higher among Muong than that among Kinh, the author concludes that social distance between Muong and Kinh seems to be bigger for Kinh than for Muong. Similar results are found by the anthropologist Dương Bích Hạnh (2008) in the context of Sa Pa in the Northern highlands between Hmong and Kinh, or by McElwee (2008) between Kinh and several minority groups, including Vân Kiju, Pa Cô, Pa Hy, and Ka Tu in the mountains of the central coast where cultural stereotypes held by both sides, Kinh migrants and minorities, continue to shape social relations. Thus, prejudice and stereotypes regarding ethnic minority people are prevalent, and the

[14] However, communes composed of a single ethnic group were an exception as different ethnic minority groups were living next to one another (Vasavakul, 2003).

media contributes to spread them, as shown by ISEE (2011) in a review of 500 articles from the four most popular newspapers.

Social distance between ethnic minorities and the majority, which may not be symmetric, could lead to the segregation of ethnic minority people from Kinh network, through inbreeding homophily process. Studying the Hmong in northern Vietnam, the anthropologist Jean Michaud (2008) shows that in Sa Pa and Bac Hà, despite the local ethnic composition where Hmong constitute the majority, not only political entities like Popular Committee are dominated by Kinh people, but also they favour Kinh entrepreneurs at least by cultural inclinations. Consequently, Hmong are excluded from Kinh economic networks.

## 3 The data and definition

### 3.1 Data source

This study uses the Vietnam Access to Resources Household Survey (VARHS).[15] The VARHS includes both commune and household questionnaires. The commune questionnaire contains demographic information on the commune, shocks, infrastructure, access to services, and development programmes. The household questionnaire covers the household roster, general characteristics of its members and housing; a module on land and crop agriculture; employment; food expenditure, savings and credit; shocks and risk coping; migration; political connections and social capital. Initially, around 2,000 rural households sampled for the 2002 and 2004 Vietnam Household Living Standards Survey (VHLSS)'s income and expenditure components were surveyed in the VARHS. In 2006, 126 randomly chosen households were supplemented to the sample in order to constitute a representative sample for rural population of 12 provinces (Brandt and Tarp, 2017). Those provinces include one province in the Red River Delta, four provinces in the North; three in the Central Coast, three in Central Highlands and one on Mekong River Delta. The data is representative at provincial level. The survey has been then implemented every two years from 2008 to 2016 in 467 rural communes; with more than 2,200 households each year and a balanced panel of 2,131 households.

We explore the modules on social network as well as the module on shocks and risk coping. We also match the VARHS with the Vietnam Census of 2009[16] to extract information on ethnic composition and occupation at the commune level. We are able to calculate the share of each ethnic group in the commune and construct an index of ethnic heterogeneity and occupational diversity at the commune level. Then, we investigate to which extent social networks differ between ethnic minorities and the Kinh majority in Vietnam's rural areas; explore how much of these differences is due to communes' ethnic compositions and to environmental contexts. For the last purpose, we acquire information from GeoQuery[17] to characterize households'

[15] This survey has been collected under the long-term collaboration between the Development Economics Research Group (DERG), University of Copenhagen and different Vietnamese institutes, namely Central Institute for Economic Management (CIEM), the Institute of Labour Science and Social Affairs (ILSSA), and the Institute of Policy and Strategy for Agriculture and Rural Development (IPSARD).

[16] This is the latest Population and Housing Census in Vietnam, conducted in 2009 by General Statistics Office with technical and financial support from the UNFPA. The census covered all Vietnamese citizens usually residing in Vietnam as of the census date and Vietnamese persons who had been permitted to go overseas within an authorized period. The 15 per cent sample of the Census is used in our study.

[17] GeoQuery is an open source providing access to custom subsets of spatial data. Data requests are available at http://geo.aiddata.org/query.

environment contexts, including physical elevation, travel distance, and night time light density. We match the Census, Geo-data and VARHS by name of communes. Due to the matching gap, we are able to match 428 communes out of 467 communes. Our study is then based on a sample of 11,180 observations in 428 communes.

### 3.2 Risk sharing networks measures

We use the name generator approach to measure risk sharing network; this approach has been widely employed in social network literature (Bidart and Charbonneau, 2012). In VARHS, the name generator tool is used to ask each household head to name the three most important contacts that he/she would ask for help in case of cash emergency;[18] then, a series of questions are raised to get detailed information on ethnicity, occupation or living area of those contacts.

The advantage of using name generator approach in social network studies is that we are able to capture broadly rich data about one household's network (ego-centered network) composition. However, we use a censored measure of social network , which could lead to bias our results: since we only have information about the three most important contacts, it is likely that the household head would think first about the ones who are easily remembered, likely the closer of his/her contacts, such as relatives or co-ethnic people. Burt (1986) identified it in network data as the 'kinship bias'. In other words, risk sharing network based on this censored information could be more homophile in general; the bridging tie or heterogeneity in their network would be underestimated accordingly. As a result, the difference in risk sharing network between Kinh and non-Kinh, due to the higher heterogeneity in Kinh's network, could be underestimated. We will take this potential bias into account in interpreting our results. Detailed discussion of the censored network bias is in Appendix (Table D). We show that the uncensored network size is indeed bigger than the censored one, by 1.8 contacts on average. The gap is positively correlated with male-headed households, households with higher educated heads, better off households, and Kinh households.

Corresponding to the research questions and hypothesis we raise, risk sharing network of the households are investigated by size and by similarity feature. Firstly, size of the household's network is measured by the total number of people that the households could reach for if they need help. This measure does not rely on the name generators but on a general question on how many people can be asked for help in case of cash emergency. Therefore, the size of the network is not censored.

Secondly, based on the name generators, we calculate the same ethnic ties proportion within the risk sharing network of each household in order to examine whether the likelihood of Kinh to link up with another Kinh is higher than their relative share in the commune. In fact, this co-ethnic share in their network is not limited within 'Kinh and non-Kinh' but corresponds to the ethnicity of the household. For instance, if the household head is Thai, we can calculate the proportion of Thai in his/her network.

Given our hypotheses that the risk sharing network of non-Kinh are more homogenous than Kinh, we measure the network's similarity by: i) the proportion of same occupation ties[19]; and ii)

[18] More specifically, the question asked is: 'If you were in need of money in case of an emergency, who outside of your household could you turn to; who would be willing to provide this assistance?' This question is widely used in the literature to capture risk-sharing networks.

[19] For each contact named, the household head is asked: 'Does [NAME] have the same occupation like yours?'. The proportion of same occupation ties is the ratio between the number of contacts for whom the household head answered yes to this question, and the total number of contacts.

the proportion of within-village contacts in their risk sharing network.[20] The underlying reason for our indicator selection is that depending on the type of shocks, these characteristics of household's risk sharing network will play a different role. To cope with a covariant shock that happened at the village scale, knowing someone from another village could benefit the household (Bramouillé and Kranton, 2007); and to cope with a covariant shock which is related to household's business and occurs on a large scale, knowing someone who has a different occupation will be an advantage (Fafchamps and Gubert, 2007).

### 3.3 Ethnic groups: definition, distribution and heterogeneity

In this study, we consider a household as ethnic minority if the head of the household belongs to any other ethnic groups than Kinh majority. Ethnic identification is one of the main concerns in studies on ethnic minorities since it is considered as 'multidimensional, fluid and contextual' (Pasquier-Doumer and Brandon, 2015). Thus, the literature provides two approaches to identify ethnicity, which are self-identification and objective ethnic markers, in which language is the most common measure used in the later concept. In Vietnam, however, official classification exists and it is widely accepted by the population. According to this official category, Kinh ethnic group which accounts for 85 per cent of the population is considered as the ethnic majority while 53 other ethnic groups are defined as ethnic minority groups in Vietnam. Thus, self-assessment is employed to identify ethnicity in administration data and surveys like the 2009 Census and VARHS. Although the Hoa/Chinese group is often combined with Kinh as the majority in several studies due to the well-off status of the Hoa (van de Walle and Gunewardena, 2000), we consider them as an ethnic minority since there are only four Hoa households in VARHS data.

We are aware of the fact that ethnic minority is not a 'homogenous' group and it would be ideal to distinguish each group. However, due to the small number of observations of each ethnic minority group, our study mainly focuses on aggregating all ethnic minority groups as one group in order to compare with Kinh majority. In an extension, we decompose the ethnic minority groups into Thai, Tay, Hmong and others. The choice is based on the fact that these groups are among the four largest ethnic minority groups in Vietnam, and thus account for a non-negligible share in our sample of ethnic minorities. Another limitation is that we are not able to take into account interethnic marriage of the heads of household that may affect the composition of their network. Unfortunately, we do not have information on ethnicity of any other household member than the household head, so we are not able to control for this effect. However, interethnic marriage are rare in Vietnam. According to our own calculation with the 2009 Census, So, the figure of inter-ethnic marriage in Vietnam (according to the Census 2009), only 4 per cent of married people have a spouse from another ethnic group.[21] The more detailed analysis of Dang and Nguyen (2015) confirms the low prevalence of inter-ethnic marriage in Vietnam, which suggests that the underestimation of the similarity in the ethnic composition of the network may be of little concern.

To explore the local patterns of ethnicity distribution, we map the data from the 2009 Census to display the geographic distribution of the Kinh, and other ethnic groups at the commune level, as presented in Figures 1 and 2. Aforementioned, Vietnam is an ethnically diverse country (Figure 1). The Kinh live almost everywhere in Vietnam but mainly occupy the delta and the coastal areas while ethnic minorities mostly live in the highland areas, especially in central and

[20] For each contact named, the household head is asked: 'Does [NAME] live in this village?'. The proportion of within-village contacts is the ratio between the number of contacts for whom the household head answered yes to this question, and the total number of contacts.

[21] For the Kinh, this proportion is 2 per cent only. It is 12.3 per cent for non-Kinh people.

northern Vietnam (Figure 2). These maps highlight the spatial segregation between ethnic groups in Vietnams' rural areas. Ethnic minorities tend to live in areas where barriers to the expansion of social networks can be important: high ethnic diversity, greater remoteness and large transport costs. Besides, as discussed above, thanks to various migration waves of the Kinh into ethnic minority areas, more than 70 per cent of the communes are mixed, and Kinh are present in almost all, but represent a minority in those mountainous and heterogeneous communes.

Figure 1. Ethnic polarization in Vietnam

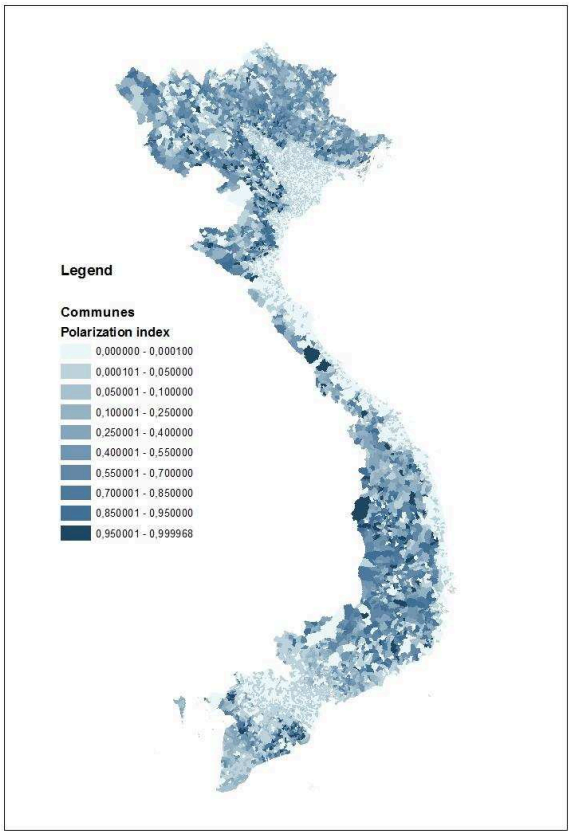
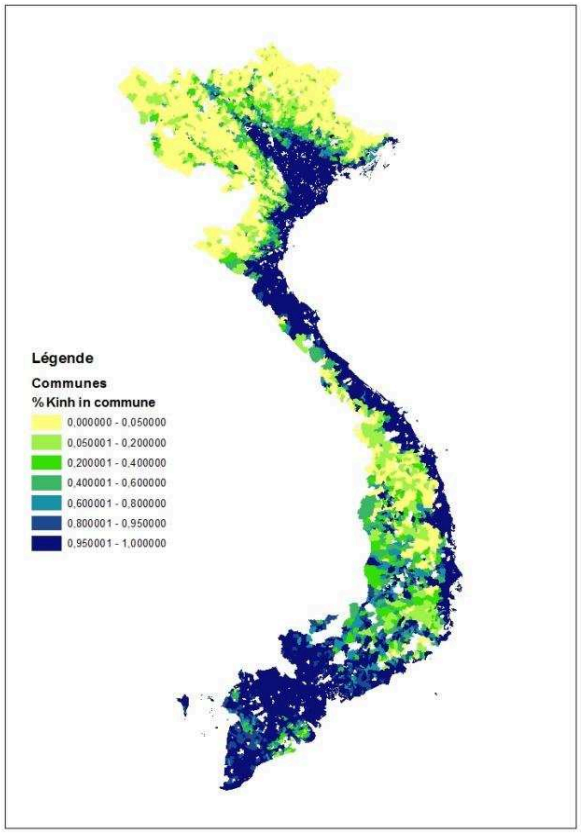



Figure 2. Kinh distribution in Vietnam




Source : 2009 Census and GeoQuery, based on authors' calculations.

In addition to ethnic distribution, we build a measure of ethnic diversity at the commune level, based on 2009 Census data. In the literature on ethnic heterogeneity studies, the two most common measures of ethnic diversity are ethnic fragmentation and ethnic polarization (Esteban and Ray, 2011). In this paper, we focus on the ethnic heterogeneity that is expected to strengthen inter-ethnic group social distance and thus ethnic inbreeding within groups. The ethnic polarization index is relevant for this purpose: theoretical and empirical studies have shown a close link between ethnic polarization and ethnic tension or conflict (Moltano and Reynal-Querol, 2005, Esteban and Ray, 2011). The main idea is that ethnic tension is higher in a situation where an ethnic majority group faces a large ethnic minority group (Horowitz, 1985), meaning ethnic conflict might occur more under the condition of a few relatively large-sized ethnic groups rather than in the case of one group facing several small-sized groups. If the ethnic tension is higher in a more polarized commune, the inter-ethnic connection between different groups would be lower, leading to a higher ethnic segregation in risk sharing network.

We use the polarization index that is originally proposed by Moltano and Reynal-Querol (2005) under the form of:

$$\text{RQindex}_c = 1 - \sum_{i=1}^{N} \left(\frac{0.5 - \pi_{ic}}{0.5}\right)^2 . \pi_{ic}$$

Where $\pi_{ic}$ is the share of ethnic group *i* in commune *c*. The Reynal-Querol polarization index (RQ index) can be interpreted as how far the distribution of ethnic groups in commune *c* is from the bimodal distribution. In other words, this index is at the highest value when there are only two groups in a commune with the same size. Figure 1 with the ethnic heterogeneity in Vietnam is based on this polarization index calculation.

### 3.4 Geographical environment/ cost of physical distance

Our objective is to understand the determinants of network formation between the ethnic majority and minorities, thus, we need to disentangle the differences in households' responses not only to their households' characteristics, particularly ethnicity factor, but also to communes' features. In Vietnam, ethnic minorities often live in geographically difficult environments (i.e. mountainous). A recurrent question in studies on ethnic gap in Vietnam is whether ethnic gap is mostly determined by unobserved differences in living areas between ethnic groups, in particular remoteness (Baulch et al., 2007, Epprecht et al., 2009, van de Walle and Gunewardena, 200). At the same time, according to the risk sharing network literature, physical distance is one barrier in forming the network. Indeed, literature on risk sharing arrangements highlights the importance of transaction costs in forming a connection, especially in developing countries because of the limitations in financial systems and infrastructure (Jack and Suri, 2014). Aida (2015) shows that due to the strong association between transaction costs embedded in monitoring or enforcement efforts and physical distance, we can use physical distance as a proxy for this type of cost in the formation of risk sharing network. Empirical results (De Weerdt, 2004; Fafchamps and Gubert, 2007) have reassured the negative role of physical distance.

Thus, controlling for the geographical disparity among ethnic groups would help us to understand not only the role of ethnic homophily but also the role of geographical remoteness in risk sharing formation.

In our study, we use four different measures of physical distance or remoteness. First, we use elevation measure, indicating the mountainous areas. Second, we calculate the difference between the minimum and maximum of elevation in each commune; this measure would help us to identify the variation of the terrain height, within each commune. Indeed, given the same mean of elevation, living in a flat surface would be easier for people to travel rather than in a high, sloped terrain. Our third measure is the average travel time between a commune and the nearest city of at least 50,000 people. Our argument for this measure of distance is that the shorter the time in commuting, the less costly for people to reach out for a relatively further connection. Traveling time also overcomes the limitation of using distance in kilometres since we cannot control for the quality of roads. Besides, easy accessibility to a relatively big town provides higher opportunities for people to get access to bigger market for agricultural products, non-farm jobs, public services such as healthcare and education (Epprecht et al., 2009), and other social-cultural activities; as a consequence, people can have more chance to diversify their network. Lastly, we use the night time light density from 1996 to 2000 as a proxy for the differences in development of each area. As recently discussed in the literature, nighttime light intensity is considered as a good proxy for economic activity and development (Henderson et al., 2012, Hodler and Raschky, 2014). In turn, travel cost are supposed be negatively associated with economic activity and development. We use all geographical data for the year 2000 and before to mitigate any reverse causality biases.

## 4 Descriptive statistics

We start our analysis with a comparison between ethnic groups of the households' characteristics, including their shock exposure, the characteristics of the risk sharing network of their head, and their geographical environment. In addition, we aim at providing a broad view of the socio-economic differences between Kinh and non-Kinh,whether they live in ethnic mixed communes or in non-mixed communes. We consider a commune as mixed if there are at least two ethnic groups living in the commune, then, there are 119 non mixed communes and 309 mixed communes. Since we are interested in the differences in network formation between Kinh and non-Kinh, we only consider mixed communes in which Kinh households are present, excluding three mixed communes without Kinh in our descriptive analysis.

Table 1 presents the differences in welfare characteristics between Kinh and non-Kinh households. It is seen that Kinh households are always better off than ethnic minorities in terms of monthly food consumption or annual income per capita; they have higher savings and also better access to formal and informal loans. Even though Kinh households are less likely to receive public transfers, the amount of their transfers is still higher than those of non-Kinh households. This result seems to support the possibility that in Vietnam, policies targeting ethnic minority are designed based on the geographical area but not directly minority households, thus, the benefits of those policies would be captured better by the Kinh living in those areas than ethnic minorities (Van de Walle and Gunewardena, 2001). It is also noticeable that across all indicators, the welfare gaps are widened in mixed communes.

**Table 1. Household's characteristics by ethnicity and commune's ethnic heterogeneity**

| | Kinh (n=8,858) | Non-Kinh (n=2,317) | Difference (T-test) | Kinh (n=5,819) | Non-Kinh (=2,159) | Difference (T-test) |
|---|---|---|---|---|---|---|
| | In all 428 communes | | | In 306 mixed communes with Kinh | | |
| Per capita food consumption (monthly) | 465.19 | 247.58 | 217.61*** | 476.04 | 249.81 | 226.24*** |
| Per capita income (annual) | 24,938.69 | 13,136.48 | 11,802.21*** | 25,803.62 | 13,354.95 | 12,448.68*** |
| Irrigated land area owned | 3,689.64 | 4,582.92 | -893.28*** | 3,308.10 | 4,669.33 | -1,361.23*** |
| Value of durables | 68,758.28 | 19,532.35 | 49,225.93 | 84,222.94 | 19,857.34 | 64,365.59 |
| Savings | 40,645.36 | 12,999.56 | 27,645.80*** | 45,027.79 | 13,350.08 | 31,677.71*** |
| Total loans | 29,384.51 | 12,039.47 | 17,345.03*** | 32,725.43 | 12,297.28 | 20,428.16*** |
| Formal loans | 21,455.01 | 8,972.24 | 12,482.77*** | 24,430.30 | 9,130.55 | 15,299.75*** |
| Informal loans | 5,924.50 | 2,044.58 | 3,879.91*** | 6,062.83 | 2,086.69 | 3,976.14*** |
| Private transfer | 6,956.57 | 3,032.85 | 3,923.72*** | 7,005.85 | 3,114.05 | 3,891.80*** |
| Public transfer | 5,900.75 | 4,595.08 | 1,305.67*** | 6,133.95 | 4,686.89 | 1,447.06*** |
| Private transfer dummy | 0.55 | 0.46 | 0.09*** | 0.55 | 0.46 | 0.09*** |
| Public transfer dummy | 0.39 | 0.75 | -0.36*** | 0.37 | 0.74 | -0.37*** |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

Table 2 presents the differences in shock exposure between ethnic minorities and majority. The VARHS data allows us to examine different types of shocks, including covariant shocks and idiosyncratic shocks. Within covariant shocks, it is also possible to distinguish between natural shocks such as floods, typhoons or pest infestation, and economic shocks such as crop price changes or input price changes. Idiosyncratic shocks include health shocks or death of the household's members, crime or theft, divorce or family dispute. In general, covariant shocks are much more prevalent than idiosyncratic shocks. In addition, ethnic minority households are more likely to be exposed to shocks than Kinh households, and this difference is driven by their

higher likelihood of facing natural shocks. Meanwhile, Kinh households reveal that they face more idiosyncratic shocks compared to ethnic minorities. In fact, health shock is the main driven factor of idiosyncratic shocks, and the higher health shocks answered by Kinh could be likely due to subjective perception of illness. It is well known that illness perception is positively correlated by wealth, which may explain that Kinh could be more likely to declare health issue even if their illness incidence is not higher than for non-Kinh.

**Table 2. Household's shock exposure by ethnicity and commune's ethnic heterogeneity**

| | Kinh (n=8,858) | Non-Kinh (n=2,317) | Difference (T-test) | Kinh (n=5,819) | Non-Kinh (=2,159) | Difference (T-test) |
|---|---|---|---|---|---|---|
| | In all 428 communes | | | In 306 mixed communes with Kinh | | |
| **Covariant shock** | 0.26 | 0.56 | -0.30*** | 0.26 | 0.44 | -0.30*** |
| Natural shock dummy | 0.24 | 0.54 | -0.30*** | 0.24 | 0.54 | -0.31*** |
| Pest shock | 0.15 | 0.41 | -0.26*** | 0.15 | 0.41 | -0.26*** |
| Economic shock dummy (covariate) | 0.04 | 0.05 | -0.01** | 0.04 | 0.05 | -0.010* |
| **Idiosyncratic shock** | 0.14 | 0.11 | 0.03*** | 0.14 | 0.11 | 0.03*** |
| Illness shock | 0.12 | 0.08 | 0.03*** | 0.11 | 0.08 | 0.03*** |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

So, descriptive statistics show that non-Kinh are worse off than Kinh and at the same time, they are more exposed to covariant shocks. These findings motivate us to continue examining another but closely related concept that is vulnerability. [22] Results are presented in Appendix 2. Two main conclusions can be driven from this vulnerability analysis: (i) vulnerability is high[23] but especially for non-Kinh households that are more severely affected by shocks, (ii) covariant shocks are a more important source of vulnerability than idiosyncratic shocks[24], knowing from Table 2 that covariant are more prevalent than idiosyncratic shocks, and that non-Kinh experience more often covariant shocks than Kinh. This justifies why we are interested in analysing differences in risk sharing networks between the two groups in terms of size and similarities among alters.

Differences in risk sharing networks of Kinh and non-Kinh are first explored in Table 3. Regarding the size of their networks, Kinh are more likely to have larger networks. Kinh households have on average 4 contacts in their network while the size is only 3.3 for non-Kinh households. Difference in network size holds true whatever mixed or non-mixed communes are considered. Then, regarding the structure of their network, we find that proportion of the same ethnic ties in household's risk sharing network is significantly higher for Kinh than for non-Kinh. This result is in line with our inbreeding homophily hypothesis according to which non-Kinh would be segregated from Kinh's network since Kinh are more likely to link with another Kinh. We can explore the similarity among contacts in the networks in terms of occupation and location. As expected, the network of Kinh is more diverse while the network of ethnic minority groups is more homophile. The proportion of contacts having the same occupation is higher for ethnic minority households: 79 per cent of ethnic minorities' risk sharing network is people

[22] We use the measure of vulnerability as expected utility (Ligon and Schechter, 2003). Vulnerability is defined as the difference between household's utility obtained from some certainty equivalent consumption and the expected utility of consumption.

[23] The estimated vulnerability in terms of food consumption is 0.785, which means on average, utility of a household is 78.5% less than the utility in a riskless situation without any inequality across all households and time periods (consumption and utility will be equal to one if the household is not vulnerable).

[24] By decomposition, we see that one third of the vulnerability is due to poverty; aggregate shock accounts for 4%, idiosyncratic shocks make up a small proportion while the unexplained shock accounts for half of the vulnerability.

doing the same jobs while the proportion is only around 50 per cent for Kinh. Regarding the location, the network is more limited within the village for ethnic minorities than for the Kinh majority. In other word, Kinh households are more likely to ask for help across villages.

Table 3. Household's risk sharing network by ethnicity

| | Kinh | Non-Kinh | Difference (T-test) | N | Kinh | Non-Kinh | Difference (T-test) | N |
|---|---|---|---|---|---|---|---|---|
| | In all 428 communes | | | | In 306 mixed communes with Kinh | | | |
| Total contacts | 3.95 | 3.30 | 0.64*** | 11,175 | 3.93 | 3.33 | 0.60*** | 7,978 |
| Same ethnic tie proportion | 0.98 | 0.91 | 0.07*** | 10,487 | 0.98 | 0.91 | 0.07*** | 7,466 |
| Relative contact proportion | 0.72 | 0.66 | 0.06*** | 10,487 | 0.73 | 0.66 | 0.07*** | 7,466 |
| Same occupation proportion | 0.51 | 0.79 | -0.28*** | 10,487 | 0.50 | 0.79 | -0.28*** | 7,466 |
| Within village contact proportion | 0.70 | 0.83 | -0.12*** | 10,487 | 0.70 | 0.82 | -0.12*** | 7,466 |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

Given our interest in the role of social distance and remoteness, we explore how social and geographical environment's factors vary between non-mixed and mixed communes in Table 4. All indicators show that geographic distances are worse in mixed communes where ethnic minority mostly live than in non-mixed communes: mixed-communes are characterized not only by high altitude mountainous areas but also by a large variance of elevation/uneven terrain, high travel time to the nearest 50,000 population city and low level of night time light. Accordingly, ethnic minority groups living in ethnically heterogeneous environment have more constraints to build a diverse network because transportation cost inside the commune is higher suggesting higher transaction cost with people out of the village. These geographical differences come along with less occupation diversity in mixed-communes as shown by the higher share of farming activities in these communes or the lower value of the index of occupational diversity.[25] This reflects lower opportunities in mixed-communes to form diverse risk sharing network in terms of occupation than in non-mixed communes.

**Table 4. Descriptive statistics of communes**

| | All communes (n=428) | | 119 non mixed communes | | 306 mixed communes | | Difference (T-test) |
|---|---|---|---|---|---|---|---|
| | Mean | (SD) | Mean | (SD) | Mean | (SD) | |
| **Ethnic composition variables** | | | | | | | |
| Share of Kinh | 0.76 | 0.37 | 0.97 | 0.18 | 0.68 | 0.39 | 0.29*** |
| Share of Tay | 0.02 | 0.09 | 0.00 | 0.00 | 0.02 | 0.10 | -0.02** |
| Share of Thai | 0.06 | 0.20 | 0.01 | 0.09 | 0.09 | 0.23 | -0.08*** |
| Share of Hmong | 0.05 | 0.18 | 0.03 | 0.16 | 0.06 | 0.19 | -0.04* |
| Ethnic polarization index | 0.20 | 0.31 | 0.00 | 0.00 | 0.28 | 0.33 | -0.28*** |
| **Geographic variables** | | | | | | | |
| Elevation | 286.30 | 374.23 | 69.75 | 199.93 | 365.23 | 391.24 | -295.48*** |
| Difference between min and max of elevation | 411.61 | 557.47 | 174.07 | 317.54 | 499.45 | 602.92 | -325.39*** |
| Distance (travelling minutes) to the nearest city in 2000 | 175.17 | 155.33 | 112.95 | 95.93 | 198.16 | 167.42 | -85.20*** |

[25] See section 3.3 for a presentation of this index.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Mean of night time light from 1996 to 2000 | 3.42 | 4.42 | 4.70 | 4.67 | 2.96 | 4.24 | 1.74*** |
| Number of villages | 10.93 | 8.04 | 9.48 | 9.94 | 11.50 | 7.14 | -2.02** |
| **Occupation variables** | | | | | | | |
| Occupational diversity index | 0.45 | 0.26 | 0.53 | 0.25 | 0.43 | 0.26 | 0.10*** |
| Farming (agriculture, forestry and fishing) | 0.69 | 0.24 | 0.65 | 0.21 | 0.71 | 0.24 | -0.06** |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

There is, even though not new, a striking difference between Kinh and non-Kinh's welfare and shock experience. Ethnic minority people are also found to be more vulnerable due to aggregate shocks, meanwhile their risk sharing networks are characterized by higher similarity, which would prevent them from coping with those shocks, according to the literature. The gaps between them stay unchanged or even larger in some indicators when we look at those who live in mixed areas. Lastly, ethnic minorities live in more diverse communities and remote areas, which are supposed to be highly correlated with low level of network diversity.

# 5 Econometric analysis

## 5.1 Empirical strategy

Section 4 revealed important and significant differences among Kinh and non-Kinh risk sharing networks, the first ones being larger, more homogenous in terms of ethnicity, but more diverse regarding other characteristics (occupation and geographic locations). Kinh households in the sample appear therefore to be better protected, thanks to their networks, against covariant risks as compared to non-Kinh.

Yet, the drivers of these differences remain to be identified. Indeed, they could result from various factors and mechanisms, which are either linked directly to ethnicity, or only indirectly. Understanding the nature of the links, and thus source of inequality, is important, in order to design appropriate policy measures.

For instance, the observed gap may be due to differences in environment that result from the spatial segregation in rural areas between Kinh, and ethnic minority populations. Rural Kinh households tend to live mostly in lowland areas, closer to urban centres, while ethnic minorities are historically settled in upland areas, which are more remote and less accessible. Different migration waves in the past decades have increased ethnic mix in both upland and lowland areas, yet, as Figure 2 shows, segregation is still strong. Infrastructures are also less developed in upland areas, partly because they are more costly, and these are also less populated areas, making it more difficult for anyone to interact with another outside his or her village. Different environment could result in different network characteristics.

Moreover, education and wealth being two important predictors of the size and the quality of networks (Fafchamps and Gubert, 2007), the Kinh, who have better access to education, and tend to occupy more often elite positions (in the political or economic sphere) than other ethnic groups, are likely to have 'better networks'.

The objective of this paper is to shed light, beyond these factors, on the other mechanisms at play directly linked to ethnicity and that drive these differences. Homophily is a well described and important mechanism in risk sharing network formation (see section 2). Individuals minimize the costs of establishing and maintaining relations by connecting with other individuals sharing the same characteristics, ethnicity for instance. As detailed in the literature review, one can consider two types of homophily. The first one is baseline homophily, and is linked to demography (or to the relative size of groups of certain characteristics) and to the opportunities to network with individuals sharing the same characteristics. The other type is inbreeding homophily and is linked to preferences to form a link with individuals sharing the same characteristics. Both mechanisms result in differentiated networks between different ethnic groups. In the first case, differences between Kinh and non-Kinh networks are explained by the fact that the first are more numerous. In the second case, individuals tend to select themselves in order to remain in homogenous networks, even when they live in highly heterogeneous environments. Disentangling these two mechanisms is important, as indeed, inequalities when the second mechanism is at play are particularly difficult to address, as they are based on some type of discrimination.

To uncover the prevalence of these factors and mechanism, we estimate a series of models explaining risk sharing network characteristics, introducing step by step various sets of covariates that account for different factors.

We start with a model of risk sharing network size that strictly accounts for household characteristics, ethnicity, as well as time and spatial fixed effects:

$$y_{ijt} = \beta_0 + \beta_1 X_{ijt} + \beta_2 E_{ij} + \delta_t + \mu_d + \vartheta_{dt} + \varepsilon_{ijt}$$

Where $y_{ijt}$ is the log of number of helpers a household *i*, from commune *j* at time *t* could rely on in case of a cash need. $X_{ijt}$ is a set of household characteristics, that includes age, sex of the head, his/her education level, occupation, and whether he/she is native of the village he/she lives in. $E_{ij}$ is a variable indicating the ethnicity of the head. In the basic form of the model, it simply indicates whether the head is Kinh or not. $\delta_t$ are year dummies[26], $\mu_d$ are district fixed effects, and $\vartheta_{dt}$, the interaction of both. $\varepsilon_{ijt}$ is the error term, and is constituted a commune-level and an household component that can be written as follow:

$$\varepsilon_{ijt} = v_{jt} + u_{ijt}$$

$E_{ij}$, the ethnic variable, captures the difference in network characteristics between ethnic groups that is not captured by other variables. Our aim is to understand what explains this difference, simply by adding successively various sets of household or commune level variables, that are likely to further explain risk sharing networks, and are correlated with the ethnic dummy. This way, we get a grasp of what triggers differences in risk sharing networks, between wealth,

[26] We pool data from all years. This allows us to use a larger set of data, to increase precision in our estimates and also to smooth measurement errors. Standard errors are clustered at both household and commune levels to account for autocorrelation. We include time and district fixed effects, but do not explore the dynamic feature of social network formation.

geography and demography. Additionally, this enables us to disentangle the mechanisms of baseline and inbreeding homophily.

We first add in the model a variable capturing household wealth ($W_{ijt}$), that is, the log of the total value of assets possessed by the household at time *t*. This variable is expected to capture households' permanent income, and to be less volatile than other conventional measures based on income or expenditure. We anticipate that household wealth is strongly correlated with both risk sharing network and ethnicity and thus that it captures part of the difference between ethnic groups. If the ethnic variable becomes insignificant, this is a sign that observed gap mostly hinges on wealth differences, or in other word, this would mean that if ethnic minorities were able to reach the same level of wealth than the Kinh, their risk sharing network would not differ significantly.

In the next model, we add commune covariates capturing the geographic context at the commune level ($G_{jt}$). In all models, district fixed effects already control to some extent for geographic conditions. Yet, the situation may vary quite strongly within districts, in terms of accessibility notably, and thus we add commune level controls (see section 3). Lagged data are used when available in order to limit risks of endogeneity, coming from the fact that social cohesion within a commune could be an unobservable omitted variable explaining both some features of risk sharing networks, and the quality of public goods or economic growth.

The following models introduce successively variables describing ethnic demography within communes. We start with a variable that measures the share of household heads' co-ethnics in the commune's population. The data is taken from the 2009 census, and is invariant in time. The introduction of this variable enables a direct test of baseline ethnic homophily, as indeed, if characteristics of the networks – namely the size or the share of co-ethnics within risk sharing networks – evolve along with the relative size of ethnic groups within communes, we can conclude that baseline homophily drives risk sharing network formation. Furthermore, if after introducing the variable, the ethnic one is no longer significant, then baseline homophily is the main mechanism at play, and inbreeding homophily is absent. But if both the share and the ethnic variables are significant, then inbreeding mechanisms may be at work.

Next, we test whether these mechanisms vary when ethnic groups are a minority or a majority within the commune, by interacting the ethnic variable with the share of co-ethnics.[27] Indeed, there is evidence from the literature that social distance between ethnic groups may strongly vary according to the relative size of each group.[28] The Kinh for instance may form different networks whether they constitute a minority in the commune's population, or whether they are the major group. And the same could be for other ethnic groups. We also test whether the number of ethnic groups in the commune and their distribution matter in the way households form risk sharing networks. Indeed, the literature on ethnic diversity and ethnic polarization suggests that not only does one's own relative size matter in the interaction between groups, but also the distribution of other groups (see Arcand and Jamovich, 2014). We introduce as explanatory variable the polarization index described in section 3.3 to capture this effect.

[27] In the basic form of the model, we test whether mechanisms vary when Kinh people are a minority or a majority in the commune by interacting the dummy of being Kinh with the share of co-ethnics of the head. In the more sophisticated model, we test it by considering four ethnic groups, Kinh, Thai, Tay and Hmong.

[28] See for instance Posner (2004) who studies the social relations between two ethnic groups that are adversaries in Malawi but allies in Tanzania. He shows that this can largely be explained by the relative size of each group within both countries' population where they are two of the main groups in the first countries, and minorities in the second.

In all of the above models, the dichotomization of the ethnic (i.e. Kinh vs. non Kinh) may appear too simplistic, especially considering that Vietnam counts 54 different ethnic groups, and that there is considerable heterogeneity between the different ethnic groups in how socially distant they are from the Kinh. We thus estimate the same model as above, but this time, the ethnic variable distinguishes five groups: the Kinh, the Thai, the Tay, the Hmong, and the rest of ethnic groups (see section 3). The first category is used as reference in all regressions. We also interact this categorical variable with the share of co-ethnics in order to see whether the different ethnic groups react similarly to Kinh when they are a minority in a commune or a majority.

The objective of this empirical strategy is to detect mechanisms at play in the formation of risk sharing networks, and in particular the existence of inbreeding homophily. We detect this last one by controlling for all possible factors that can capture variation between ethnic groups, the remaining effect in the ethnic variable is interpreted as inbreeding homophily, i.e. as ethnically based preference. Yet, if an omitted factor, which is not preference but is correlated with both risk sharing network characteristics and ethnicity, remain in the error term, the coefficient of the ethnic variable is likely to be biased. We thus estimate the model adding commune-level fixed effects, and removing all commune-level variables, to check whether our results can be explained by such omitted factors.

We estimate similar models to explain four characteristics of risk sharing networks described in section 4. The first is network size, models are estimated using log-linear regressions. The second is the share of co-ethnics within risk sharing networks, estimated through tobit regressions, then come models of geographic similarity, measured by the share of network members living in the same village as the head, and then models of occupational similarity, i.e. the share of network members having the same occupation as the head.

## 5.2 Results

*5.2.1 Determinants of household network size*

Results from the OLS estimates are shown in Table 5.[29] As expected, we observe that male and more educated household heads have larger networks.[30] We also find a positive and significant effect of the Kinh dummy. The larger size of Kinh's network is observed when controlling for the characteristics of the head (sex, age, education, farm activity, place of birth), the size of the household, the time of the survey, and the time-invariant unobserved characteristics of the district (column (1)). However, the Kinh dummy becomes insignificant as far as the wealth of household is controlled for (columns (2)-(6)), indicating some collinearity between both variables. Thus, the Kinh's 'advantage' in network size is fully captured by wealth.

In addition, we find that remoteness and the network size are positively associated, reflecting the barrier of physical distance in forming risk sharing network. However, there is no effect of ethnic composition on network size. This result remains stable over various specifications of the composition variables: the network size is not associated with the relative size of the ethnic group of the household head (column (4)), with the ethnic dummy (column (5)), and it does not depend on how ethnic groups are distributed within the commune (column 6)).

The results thus indicate that Kinh households have larger risk sharing networks, mostly because they tend to be richer than other groups within their districts, and not because they are more numerous. This result generalizes Fischer et al. (2010)'s findings. Their study analyses the social network of 33 ethnic minority individuals living in the uplands of Northern Vietnam. It shows that household's level of wealth is an influential factor with regard to the formation and size of the network, mostly because of the cost of maintaining relationships. Poorer households purposely refrain from asking for help from their sparse networks because they fear being unable to reciprocate later.

However, once we use the commune fixed effect in the last model (column (7)), we again find a significant effect of Kinh dummy. This result indicates a higher preference for large networks among the Kinh. This result could be due to the fact that some unobservable characteristics of communes that we do not capture well with explanatory variables, are highly correlated with Kinh. We could think about customs or quality of institutions of the communes supporting connections between households that are more present in some areas where Kinh households live. Therefore, those unobserved characteristics of the communes lead to the difference in our commune fixed effect and district fixed effect models.

[29] Results from the Poisson regressions are available upon request from the authors.

[30] We consider the network described in the data to be the individual network of the household heads, who is the respondent in most cases. We consider that relations of other household members were not taken into account.

*5.2.2 Determinants of ethnic similarity within risk sharing networks*

We then estimate the determinants of ethnic similarity. Results of the tobit estimates are reported in Table 6.[31] Contrary to network size, household head characteristics do not affect significantly the degree of ethnic similarity within risk sharing networks: the level of ethnic similarity within the risk sharing networks is not affected by head characteristics such as age, household size, education or gender. Farmers tend to have more homogenous networks than heads with other occupations.

We find a positive and significant effect of the ethnic variable, indicating that the Kinh are more likely to network with other Kinh people than the non-Kinh to form links with people sharing the same ethnic group. This result is not changed after controlling for household wealth, which does not significantly explain ethnic homophily (column (2)), or after controlling for a potential censoring bias in the dependent variable (column (3)). We test and control for such bias by introducing a dummy variable indicating whether the household indicated having more than three contacts whom to ask help in case of need, and another variable reporting the total number of potential helpers. None of these variables have significant coefficients, unless in the commune fixed effect models.

Differences in geography do not seem to be a source of difference between Kinh and non Kinh risk sharing networks (column (4)). The introduction of variables capturing this factor does only slightly affect the coefficient on the ethnic variable, but this last remains strongly significant and positive.

The model displayed in column (5) includes the share of household head's co-ethnics in the commune among regressors. Its coefficient is positive, significant and of high magnitude. Furthermore, it captures part of the ethnic effect, whose coefficient is almost halved, yet still positive and significant. This result indicates that baseline homophily is at play in the formation of risk sharing networks: the more co-ethnics one has in the commune, the more he or she has in his or her network. Kinh risk sharing networks are thus homogenous partly because they live in communes with a higher share of co-ethnics.[32] The coefficients obtained after interacting the share of co-ethnics with the Kinh dummy variable (column (6)) indicates that the effect of the relative ethnic size is positive for both non-Kinh and Kinh households, but much stronger for the later. The Kinh variable alone, however, is negative, but non-significant.

We observe that this mechanism does not capture all of the differences between ethnic groups, and the positive and significant coefficient of the ethnic variable indicate some inbreeding mechanism, i.e. a selection of co-ethnics in the risk sharing network that is based on preferences. The Kinh appear to be more selective than other groups. The decomposition of the ethnic variable between different ethnic groups as shown in Table E in the Appendix (column (3)) shows that Kinh are indeed more selective than the Thai and the Tay, but do not differ significantly from the Hmong or other ethnic groups on average.

[31] We display here the tobit estimates that appear to be the best specifications given the distribution of the dependent variable. The variable of the share of same ethnic members is indeed highly skewed with a value equal to 1 in 95% of cases. OLS results are available upon request to the authors.

[32] We tested in models whose results are not shown here, the linearity of this effect, by introducing the square of the 'share' variable, or by introducing it as categorical variable taking difference levels (less than a 25%, from 25% to 50%; from 50 to 75% from 75% to 99%, and from 99% to 100%). The objective was to test whether there is a specific effect of being a minority vs. a majority, or if there is a specific effect for homogenous communes. Results show that the degree of ethnic homophily increases with the size of the share, proportionally to this share, and that the relation between both is linear.

Taken all together these results indicate that Kinh are more likely than others to link up with people from their ethnic group, particularly when they live in communes where their ethnic group represents a significant share of the population. In addition, baseline homophily or demographics of ethnic groups is not enough to explain the likelihood to form link with co-ethnics. Inbreeding homophily is at play and is stronger for Kinh people: they are more likely to network with co-ethnics once the demographics of the ethnic groups is controlled for; and the more in majority they are, the more inbreeding homophily they experience.

Finally, we find that ethnic similarity in risk sharing networks is negatively influenced by the level of ethnic polarization, often used to test for the potential conflicts that may emerge when several groups of similar size coexist within a same commune. We would expect if such conflictual situation existed in sampled communes, that a higher ethnic polarization would pull population apart, and accentuate segregation within networks. However, the negative and significant sign of the polarization index variable shows that quite the opposite occurs: the more the commune is polarized, the lower the inbreeding homophily. In fact, among 23 communes where the polarization index is higher than 0.9, we find 21 communes where the Kinh share a comparable size with another ethnic group such as Thai, Tay, Ede or Co Ho. These groups are amongst the well-off minority groups with relatively high level of education (MRDI, 2014), thus, it is possible that the social distance between these groups and the Kinh is shortened, facilitating ethnic integration.

Last, we estimate the model with commune-fixed effects controlling for all observable and unobservable characteristics at this level. The ethnic dummy coefficient remains positive and significant but is of lower magnitude, and the share of co-ethnics also remains high, positive and significant. This result confirms a baseline homophily mechanism, as well as a tendency for inbreeding homophily among the Kinh who have higher preference than others to remain in ethnically homogenous networks.

*5.2.3 Determinants of geographic similarity within risk sharing networks*

Another particularly important feature of a risk sharing network lies in its geographic dispersion. A high proportion of the shocks faced by populations in rural areas are related to climate, and affect whole populations within a given areas. Thus, one's capacity to rely on someone located outside its immediate environment, when such a shock occurs will affect its ability to face it. The literature from the new economics of migration has well shown for instance that the departure of some family members to further places is partly driven by the necessity to insure the rest of the family. Also, a recent study by Jaimovich (2014) highlights the importance of external economic links which help to connect missing markets in the isolated communities.

We look here at the determinants of geographic proximity of risk sharing network members, measured by the share of the network members who live in the same village as the household head. Results of the tobit estimates are reported in Table 7. We find that more educated and richer households are more likely to have connections outside the village. In the same way, farmers are more likely to have village neighbours as network members. Neither the censor variable nor the size of the network has a significant effect, indicating that the censoring effect is not at play when it comes to geographic similarity. Regarding geographical environment, we find that living in a more developed area helps people to form network across village boundaries.

The Kinh networks are more geographically spread than those of the non-Kinh, as indicated in descriptive statistic and in the first estimation (column (1)). Risk sharing network from ethnic minority people are more likely to be confined within village boundaries, making ethnic minorities highly vulnerable to aggregated shocks. This result confirms what Fischer et al. (2010)

find based on case studies of ethnic minority people in North Vietnam. They show that the majority of help flows remain within the village. The rare cases where they go beyond the village boundary, the help flows come from the women's village of origin.

However, the Kinh dummy becomes insignificant once we control for wealth (columns (2)-(8)). Similar to network size, the advantage of Kinh people in forming distant links is highly correlated with wealth. Likewise, we do not find any effect of communes' ethnic composition on the geographic proximity of risk sharing networks. Here again, there is no evidence that a high ethnic polarization would pull individuals apart and push them to remain within their village, but also no evidence that it helps to create more connections outside one's village.

A limitation in this study is that we have no information on the ethnic composition at the village-level. It may well be that even in very heterogeneous communes, villages remain completely homogenous ethnically. However, the fact that there is a positive effect of polarization on the ethnic diversity within networks, but a non-significant effect on the geographic similarity, is an indication that when there is ethnic diversity at the commune level, there is also, to some extent, diversity within the villages.

In the last model, with commune fixed effect, our results still hold, indicating that the ability to form geographically distant links is mostly correlated with education and wealth, rather than being a majority or minority. Similar results are found by distinguishing ethnic minorities groups (Appendix Table E).[33] Kinh households have a more efficient risk sharing network in that aspect because they are usually endowed with a higher education level and they are richer than ethnic minorities. Thus, despite being more numerous and spread in the whole territory, Kinh people are not more able to form and maintain links out of the village than other ethnic groups once their socio-economic characteristics are taken into account. Baseline and inbreeding homophily are therefore not determinant mechanisms of geographic similarity in risk sharing network.

*5.2.4 Determinants of occupational similarity*

Last, we look at the occupational similarity within risk sharing networks**.** As pointed out in section 4, one essential feature of risk sharing network is its ability to protect one from the adverse effects of covariant shocks. While idiosyncratic risks can more easily be covered by homogenous networks, the insurance against covariant shocks requires being able to rely on someone that is not exposed to the same risks when such shock occurs. Occupational similarity is in this perspective an important characteristic of risk sharing networks, particularly in rural areas where incomes and consumption fluctuate strongly within a year, and where covariant shocks dominate. Occupational similarity is defined by the proportion of people in the risk sharing network having the same occupation than the household's head.[34] Estimation strategy is the same than for geographic similarity, except that we control from column (5) by the occupational diversity in the commune. Indeed, the diversity of occupations in one risk sharing network should highly depend on the local labour market and the diversity of occupations found in the commune. In order to isolate the effect of the distribution of ethnic groups and their social distance from the differences in local labour markets, we built an index of occupational diversity in the communes using 2009 census data on occupation of all adults in the commune in the last seven days. This index has to be interpreted as the probability that two people taken

[33] Only a small difference is found between Kinh and Hmong, Hmong people being significantly less likely than Kinh to form link out of the village, everything else being equal.

[34] Because this measure is based on the name generators, the proportion is calculated among the three first contacts mentioned by the household head.

randomly in a commune have different occupations.[35] Therefore, the higher the index, the larger the spectrum of occupations within the commune.

Table 8 provides the results from the tobit regressions of the occupational similarity equations. We find that Kinh variable is negative and significant in all specifications. Kinh individuals have more occupational diversity within their risk sharing networks than individuals from other ethnic groups, everything else being equal. This result cannot be explained by lower cost for Kinh people to meet with distant people, who are potentially more diverse than neighbours: the coefficient of Kinh variable is almost unchanged when proxies for the remoteness of the commune are introduced (column (4)).

Kinh people do not have more diversity in their network than non-Kinh because they live in communes where the labour market is more diversified: the effect of Kinh variable decreases only slightly when occupational diversity within the commune is controlled for (column (5)). Thus, an ethnic minority farmer is less likely than a Kinh farmer to form link with someone who is not farmer, even if non-farm activity is as frequent in his/her community than in the Kinh one, and even if both of them live in a remote commune. In addition, Kinh people are more likely to have occupational diversity with their network than minority people, even when ethnic minority groups are considered separately (Appendix Table E).

We find a positive and significant association between occupational similarity of the network and the share of co-ethnics in the commune (column (6)). But more importantly, controlling by the demography of ethnic groups increases the effect of being Kinh. It suggests that differences between Kinh and non-Kinh change depending on whether they are in majority or in minority in the commune, as confirmed by the interaction of the Kinh dummy with the share of co-ethnics in the community (column (7)). Kinh people distinguish themselves more when they are in minority: when belonging to a minority within the commune, the Kinh are more able than the non-Kinh to diversify their networks. Reversely, a Kinh in majority is not so much different (but still he/she is) from a non-Kinh in majority in that aspect. Besides, introducing the index of polarization almost does not change the effect of the ethnic variable.

Three interpretations can be drawn from these results. First, Kinh are more likely to have diverse network in terms of occupation because they are more able to form and maintain links out of the village, and ties out of the village are supposed to connect more frequently with people having different occupation. However, results on geographic similarity go against this first interpretation. They show that Kinh people are not more able to form and maintain links out of the village than other ethnic groups once their socio-economic characteristics are taken into account. Thus, according to this interpretation, Kinh variables should not be significant once socio-economic characteristics are taken into account.

The second interpretation is that Kinh people are more likely to occupy elite positions, and as such to play the role of structural hole (Burt, 1995). In other words, they are more likely to have positional advantages in the commune by connecting individuals or groups of individuals that are connected only through their intermediary. They are therefore connected with more diverse people whatever their ethnic group. But this interpretation can hold true only if socio-economic

[35] More precisely, this index is built as a fragmentation index calculated as below:

$$\text{FRAG}_c = 1 - \sum_{i=1}^{N} \pi_{ic}^2$$

Where $\pi_{ic}$ is the proportion of people with occupation *i* in a commune *c* and *N* is the total of number of occupation in the commune. Occupation is defined here as the 2-digit code of the 2007 Vietnam standard industrial classification.

characteristics introduced in the model do not capture elite position. However, we believe that variables measuring education, the value of assets, and the dummy indicating whether the head is a farmer account for this positional effect. Thus if it was the main effect, these variables should absorb most of the Kinh variable effect in the regression.

A third interpretation is that Kinh people occupy a larger spectrum of occupations than non-Kinh people, and in addition, that they are more likely to link with other Kinh, excluding ethnic minorities from their network. Consequently, the people with whom ethnic minorities can form link have a reduced set of occupations. We have shown that Kinh are more likely to link with other Kinh through inbreeding homophily process (section 5.2.2), and that ethnic minorities have restrained opportunities to link with Kinh people. In order to see whether occupational diversity within a commune is higher among Kinh than among other ethnic groups, we calculate the index of occupational diversity among Kinh people in a commune and compare it with the index of occupational diversity among other ethnic groups. We find that considering only Kinh people within a rural commune and selecting randomly two of them, the probability that they have different occupation is 0.524 (sd 0.243). Doing the same simulation for non-Kinh people, the probability is significantly lower, at 0.167 (sd 0.211). The probabilities are almost the same when we consider only mixed communes with both Kinh and non-Kinh. Therefore, we conclude to a higher occupational diversity among Kinh people at the commune level compared to ethnic minorities.

Other interpretations related to unobserved commune characteristics should be excluded as the ethnic variable is still significant when commune fixed effects are introduced (column (11)). From these results, we conclude that inbreeding homophily combined with higher diversity in occupations within Kinh people seems to be an important driver of differences between majority and minority ethnic groups regarding occupational similarity of risk sharing network.

As far as other independent variables are concerned, we find that occupational similarity within networks is weaker for the more educated individuals, and for richer households whatever the specification of the model. The negative wealth effect on occupational similarity is expected, as we suppose that building and maintaining bridging connections requires more resources than building and maintaining connection with similar people. Occupational similarity is much higher however for farmers. This is not surprising as farming is the main occupation of about 70 per cent of household heads in the sample. The probability for a farmer to link up with another farmer is therefore higher than for any other person with another occupation to link up with someone with the same occupation. The effect disappears as far as the index of occupational diversity in the commune is introduced. Surprisingly, being male is associated with more similar network in terms of occupation. Finally, we find that heads with greater network size exhibit a higher level of occupational similarity in their networks. This is likely to come from the censoring bias that affects network composition variables of household heads with large networks. The positive and significant effect is thus explained by the fact that individuals with many connections tend to cite first contacts with the same occupation than other contacts (Appendix 3). It could well be indeed, that because they share the same occupation, they are in closer and more frequent contact, and are therefore more directly thought of when the question of contacts is asked.

**Table 5. Determinants of network size – Log of number of potential helpers**

| Log of potential helpers | (1) OLS | | (2) OLS | | (3) OLS | | (4) OLS | | (5) OLS | | (6) OLS | | (7) OLS | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Household head characteristics** | | | | | | | | | | | | | | |
| Age of HH head | -0.003 | (0.00) | -0.006** | (0.00) | -0.006** | (0.00) | -0.006** | (0.00) | -0.006** | (0.00) | -0.006** | (0.00) | -0.004 | (0.00) |
| (Age of head)$^2$ | 0.000 | (0.00) | 0.000** | (0.00) | 0.000** | (0.00) | 0.000** | (0.00) | 0.000** | (0.00) | 0.000** | (0.00) | 0.000 | (0.00) |
| Head is male | 0.060*** | (0.02) | 0.048*** | (0.02) | 0.049*** | (0.02) | 0.049*** | (0.02) | 0.049*** | (0.02) | 0.048*** | (0.02) | 0.051*** | (0.02) |
| Head completed primary school | 0.062*** | (0.02) | 0.041* | (0.02) | 0.039* | (0.02) | 0.039* | (0.02) | 0.038* | (0.02) | 0.039* | (0.02) | 0.033 | (0.02) |
| Head completed lower secondary | 0.099*** | (0.02) | 0.064*** | (0.02) | 0.061** | (0.02) | 0.061** | (0.02) | 0.061** | (0.02) | 0.062** | (0.03) | 0.058** | (0.02) |
| Head completed upper secondary | 0.150*** | (0.03) | 0.106*** | (0.03) | 0.103*** | (0.03) | 0.103*** | (0.03) | 0.103*** | (0.03) | 0.103*** | (0.03) | 0.095*** | (0.03) |
| Head is farmer | -0.012 | (0.02) | -0.007 | (0.02) | -0.006 | (0.02) | -0.006 | (0.02) | -0.006 | (0.02) | -0.005 | (0.02) | 0.002 | (0.02) |
| HH size | 0.005 | (0.00) | -0.005 | (0.004) | -0.005 | (0.00) | -0.004 | (0.00) | -0.004 | (0.00) | -0.004 | (0.00) | -0.002 | (0.00) |
| Head or Spouse is village native | 0.019 | (0.03) | 0.021 | (0.03) | 0.015 | (0.03) | 0.016 | (0.03) | 0.017 | (0.03) | 0.016 | (0.03) | 0.000 | (0.03) |
| Asset value (log) | | | 0.029*** | (0.00) | 0.029*** | (0.00) | 0.029*** | (0.00) | 0.029*** | (0.00) | 0.029*** | (0.00) | 0.025*** | (0.00) |
| **Geography** | | | | | | | | | | | | | | |
| Elevation | | | | | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | | |
| Difference between min and max of elevation | | | | | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | | |
| Distance to nearest city (in 2000) | | | | | -0.000** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) | | |
| Night light from 1996 to 2000 | | | | | -0.004 | (0.00) | -0.003 | (0.00) | -0.003 | (0.00) | -0.004 | (0.01) | | |
| **Ethnicity** | | | | | | | | | | | | | | |
| Kinh | 0.067* | (0.04) | 0.052 | (0.04) | 0.040 | (0.036) | 0.048 | (0.04) | 0.050 | (0.07) | 0.051 | (0.04) | 0.097** | (0.04) |
| Kinh x share of co-ethnic | | | | | | | | | -0.004 | (0.11) | | | | |
| Share of co-ethnic in commune | | | | | | | -0.028 | (0.04) | -0.027 | (0.06) | -0.048 | (0.05) | -0.010 | (0.06) |
| Ethnic polarization index | | | | | | | | | | | -0.044 | (0.05) | | |
| Constant | 0.913*** | (0.11) | 0.775*** | (0.11) | 0.857*** | (0.117) | 0.863*** | (0.12) | 0.863*** | (0.12) | 0.905*** | (0.13) | 0.674*** | (0.10) |
| Observations | 11175 | | 11175 | | 11175 | | 11175 | | 11175 | | 11175 | | 11175 | |
| R-squared | 0.102 | | 0.107 | | 0.107 | | 0.107 | | 0.107 | | 0.107 | | 0.147 | |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects, the last column includes commune fixed effects. Source: Authors' calculations based on the 2008-2016 VARHS.

**Table 6. Determinants of ethnic similarity within risk sharing network – Share of co-ethnics among networks' members**

| Ethnic similarity | (1) Tobit | | (2) Tobit | | (3) Tobit | | (4) Tobit | | (5) Tobit | | (6) Tobit | | (7) Tobit | | (8) Tobit | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Household head characteristics** | | | | | | | | | | | | | | | | |
| Age of HH head | -0.017 | (0.04) | -0.021 | (0.04) | -0.020 | (0.04) | -0.025 | (0.04) | -0.034 | (0.04) | -0.031 | (0.04) | -0.036 | (0.04) | -0.038 | (0.04) |
| (Age of head)2 | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) |
| Head is male | 0.248 | (0.22) | 0.216 | (0.22) | 0.200 | (0.22) | 0.207 | (0.22) | 0.127 | (0.22) | 0.116 | (0.23) | 0.104 | (0.22) | 0.149 | (0.22) |
| Head completed primary school | -0.063 | (0.30) | -0.101 | (0.30) | -0.109 | (0.30) | -0.143 | (0.30) | -0.130 | (0.31) | -0.099 | (0.31) | -0.139 | (0.31) | -0.052 | (0.28) |
| Head completed lower secondary | -0.137 | (0.31) | -0.210 | (0.32) | -0.213 | (0.32) | -0.247 | (0.32) | -0.215 | (0.33) | -0.194 | (0.33) | -0.207 | (0.33) | -0.054 | (0.30) |
| Head completed upper secondary | -0.057 | (0.37) | -0.152 | (0.37) | -0.162 | (0.38) | -0.214 | (0.38) | -0.179 | (0.39) | -0.123 | (0.38) | -0.153 | (0.39) | -0.091 | (0.34) |
| Head is farmer | 0.324* | (0.17) | 0.334* | (0.17) | 0.336* | (0.17) | 0.372** | (0.17) | 0.278* | (0.17) | 0.272 | (0.17) | 0.290* | (0.17) | 0.163 | (0.19) |
| Head or Spouse is village native | 0.266 | (0.31) | 0.281 | (0.31) | 0.287 | (0.31) | 0.186 | (0.32) | -0.019 | (0.31) | -0.021 | (0.31) | -0.033 | (0.31) | 0.035 | (0.27) |
| HH size | -0.060 | (0.05) | -0.074 | (0.05) | -0.076 | (0.05) | -0.086* | (0.05) | -0.090* | (0.05) | -0.090* | (0.05) | -0.089* | (0.05) | -0.123** | (0.05) |
| Asset value (log) | | | 0.059 | (0.04) | 0.053 | (0.04) | 0.054 | (0.04) | 0.069* | (0.04) | 0.070* | (0.04) | 0.068* | (0.04) | 0.066* | (0.04) |
| Censor | | | | | 0.434 | (0.32) | 0.421 | (0.32) | 0.508 | (0.31) | 0.508 | (0.31) | 0.499 | (0.31) | 0.562** | (0.28) |
| Network size (log) | | | | | -0.070 | (0.23) | -0.078 | (0.23) | -0.113 | (0.23) | -0.104 | (0.23) | -0.107 | (0.23) | -0.125 | (0.22) |
| **Geography** | | | | | | | | | | | | | | | | |
| Elevation | | | | | | | 0.002 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | | |
| Difference between min and max of elv. | | | | | | | 0.000 | (0.00) | 0.001 | (0.00) | 0.000 | (0.00) | 0.001 | (0.00) | | |
| Distance to nearest city (in 2000) | | | | | | | -0.004*** | (0.00) | -0.005*** | (0.00) | -0.004*** | (0.00) | -0.005*** | (0.00) | | |
| Night light from 1996 to 2000 | | | | | | | 0.081 | (0.06) | 0.065 | (0.06) | 0.036 | (0.06) | 0.044 | (0.06) | | |
| **Ethnicity** | | | | | | | | | | | | | | | | |
| Kinh | 1.966*** | (0.53) | 1.952*** | (0.52) | 1.914*** | (0.52) | 1.738*** | (0.52) | 0.911** | (0.463) | -0.511 | (0.66) | 0.920** | (0.46) | 0.770* | (0.40) |
| 1. Kinh x share | | | | | | | | | | | 2.896*** | (1.11) | | | | |
| Share of co-ethnic in commune | | | | | | | | | 2.736*** | (0.476) | 1.688*** | (0.59) | 2.420*** | (0.50) | 2.987*** | (0.59) |
| Ethnic polarization index | | | | | | | | | | | | | -1.094** | (0.48) | | |
| Constant | 5.998*** | (1.44) | 5.687*** | (1.46) | 5.738*** | (1.44) | 6.163*** | (1.52) | 5.890*** | (1.600) | 5.992*** | (1.64) | 6.985*** | (1.71) | 6.633*** | (1.74) |
| Observations | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | |
| Pseudo $R^2$ | 0.154 | | 0.154 | | 0.155 | | 0.158 | | 0.171 | | 0.173 | | 0.172 | | 0.242 | |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects, the last column includes commune fixed effects.

Source: Authors' calculations based on the 2008-2016 VARHS.

**Table 7. Determinants of geographical similarity in risk sharing network – Share of networks' members living in the same village**

| | (1) | | (2) | | (3) | | (4) | | (5) | | (6) | | (7) | | (8) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Geographical similarity | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | |
| **Household head characteristics** | | | | | | | | | | | | | | | | |
| Age of HH head | -0.014 | (0.01) | -0.007 | (0.01) | -0.007 | (0.01) | -0.008 | (0.01) | -0.008 | (0.01) | -0.008 | (0.01) | -0.007 | (0.01) | -0.004 | (0.01) |
| (Age of head)2 | 0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | 0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) |
| Head is male | 0.123* | (0.07) | 0.156** | (0.07) | 0.155** | (0.07) | 0.152** | (0.07) | 0.153** | (0.07) | 0.153** | (0.07) | 0.155** | (0.07) | 0.139** | (0.06) |
| Head completed primary school | -0.144 | (0.11) | -0.086 | (0.11) | -0.088 | (0.11) | -0.086 | (0.11) | -0.087 | (0.12) | -0.086 | (0.11) | -0.088 | (0.11) | -0.065 | (0.10) |
| Head completed lower secondary | -0.316*** | (0.11) | -0.216* | (0.11) | -0.218* | (0.11) | -0.215* | (0.11) | -0.216* | (0.11) | -0.216* | (0.11) | -0.219* | (0.11) | -0.173* | (0.10) |
| Head completed upper secondary | -0.630*** | (0.13) | -0.498*** | (0.13) | -0.502*** | (0.13) | -0.501*** | (0.13) | -0.502*** | (0.13) | -0.502*** | (0.13) | -0.505*** | (0.13) | -0.438*** | (0.11) |
| Head is farmer | 0.181*** | (0.06) | 0.168*** | (0.06) | 0.169*** | (0.06) | 0.166*** | (0.06) | 0.167*** | (0.06) | 0.167*** | (0.06) | 0.166*** | (0.06) | 0.161*** | (0.06) |
| Head or Spouse is village native | 0.176** | (0.09) | 0.168* | (0.09) | 0.167* | (0.09) | 0.164* | (0.09) | 0.168* | (0.09) | 0.168* | (0.09) | 0.168* | (0.09) | 0.225** | (0.10) |
| HH size | 0.003 | (0.02) | 0.032* | (0.02) | 0.032* | (0.02) | 0.031* | (0.02) | 0.031* | (0.02) | 0.031* | (0.02) | 0.031* | (0.02) | 0.027* | (0.02) |
| Asset value (log) | | | -0.091*** | (0.02) | -0.092*** | (0.02) | -0.091*** | (0.02) | -0.091*** | (0.02) | -0.091*** | (0.02) | -0.091*** | (0.02) | -0.098*** | (0.02) |
| Censor | | | | | -0.040 | (0.09) | -0.037 | (0.09) | -0.038 | (0.09) | -0.038 | (0.09) | -0.037 | (0.09) | -0.006 | (0.08) |
| Network size (log) | | | | | 0.050 | (0.07) | 0.047 | (0.07) | 0.048 | (0.07) | 0.048 | (0.07) | 0.047 | (0.07) | 0.021 | (0.07) |
| **Geography** | | | | | | | | | | | | | | | | |
| Elevation | | | | | | | 0.001 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | 0.001* | (0.00) | | |
| Difference between min and max of elv. | | | | | | | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | | |
| Distance to nearest city (in 2000) | | | | | | | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | | |
| Night light from 1996 to 2000 | | | | | | | -0.027* | (0.02) | -0.027* | (0.02) | -0.027* | (0.02) | -0.025* | (0.02) | | |
| Number of villages | | | | | | | 0.003 | (0.00) | 0.003 | (0.00) | 0.002 | (0.00) | 0.002 | (0.00) | | |
| **Ethnicity** | | | | | | | | | | | | | | | | |
| Kinh | -0.260* | (0.14) | -0.219 | (0.14) | -0.220 | (0.14) | -0.202 | (0.14) | -0.176 | (0.16) | -0.204 | (0.29) | -0.190 | (0.16) | -0.151 | (0.194) |
| 1. Kinh x share | | | | | | | | | | | 0.055 | (0.44) | | | | |
| Share of co-ethnic in commune | | | | | | | | | -0.093 | (0.18) | -0.116 | (0.25) | 0.009 | (0.20) | -0.190 | (0.254) |
| Ethnic polarization index | | | | | | | | | | | | | 0.216 | (0.21) | | |
| Constant | 3.272*** | (0.40) | 3.756*** | (0.42) | 3.720*** | (0.43) | 3.798*** | (0.46) | 3.821*** | (0.46) | 3.825*** | (0.46) | 3.625*** | (0.51) | 3.510*** | (0.58) |
| Observations | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | |
| Pseudo $R^2$ | 0.0408 | | 0.0430 | | 0.0430 | | 0.0434 | | 0.0434 | | 0.0434 | | 0.0435 | | 0.0712 | |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects, the last column includes commune fixed effects.

Source: Authors' calculations based on the 2008-2016 VARHS.

**Table 8. Determinants of occupational similarity – Share of networks' members having the same occupation**

| | (1) | | (2) | | (3) | | (4) | | (5) | | (6) | | (7) | | (8) | | (9) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occupational similarity | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | | Tobit | |
| **Household head characteristics** | | | | | | | | | | | | | | | | | | |
| Age of HH head | 0.032* | (0.02) | 0.041** | (0.02) | 0.042** | (0.02) | 0.041** | (0.02) | 0.022 | (0.02) | 0.019 | (0.02) | 0.020 | (0.02) | 0.019 | (0.02) | 0.017 | (0.02) |
| (Age of head)2 | -0.001*** | (0.00) | -0.001*** | (0.00) | -0.001*** | (0.00) | -0.001*** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) |
| Head is male | 0.414*** | (0.09) | 0.454*** | (0.09) | 0.441*** | (0.09) | 0.430*** | (0.09) | 0.414*** | (0.09) | 0.405*** | (0.09) | 0.405*** | (0.09) | 0.402*** | (0.09) | 0.399*** | (0.09) |
| Head completed primary school | -0.094 | (0.15) | -0.025 | (0.15) | -0.034 | (0.15) | -0.014 | (0.15) | -0.005 | (0.15) | 0.005 | (0.15) | 0.012 | (0.15) | 0.006 | (0.15) | -0.029 | (0.13) |
| Head completed lower secondary | -0.382** | (0.15) | -0.257* | (0.16) | -0.271* | (0.16) | -0.249 | (0.16) | -0.236 | (0.15) | -0.222 | (0.15) | -0.213 | (0.15) | -0.217 | (0.15) | -0.215 | (0.14) |
| Head completed upper secondary | -1.066*** | (0.17) | -0.902*** | (0.17) | -0.930*** | (0.17) | -0.909*** | (0.17) | -0.875*** | (0.16) | -0.863*** | (0.16) | -0.853*** | (0.16) | -0.858*** | (0.16) | -0.792*** | (0.15) |
| Head is farmer | 0.679*** | (0.09) | 0.664*** | (0.09) | 0.671*** | (0.09) | 0.662*** | (0.09) | 0.013 | (0.12) | 0.009 | (0.12) | 0.009 | (0.12) | 0.011 | (0.12) | -0.008 | (0.11) |
| Head or Spouse is village native | -0.052 | (0.13) | -0.063 | (0.13) | -0.065 | (0.13) | -0.055 | (0.13) | -0.020 | (0.13) | -0.055 | (0.13) | -0.055 | (0.13) | -0.056 | (0.13) | -0.039 | (0.12) |
| HH size | 0.112*** | (0.02) | 0.148*** | (0.02) | 0.149*** | (0.02) | 0.147*** | (0.02) | 0.149*** | (0.02) | 0.149*** | (0.02) | 0.149*** | (0.02) | 0.149*** | (0.02) | 0.147*** | (0.02) |
| Asset value (log) | | | -0.112*** | (0.02) | -0.121*** | (0.02) | -0.118*** | (0.021) | -0.111*** | (0.02) | -0.109*** | (0.02) | -0.110*** | (0.02) | -0.110*** | (0.02) | -0.113*** | (0.02) |
| Censor | | | | | 0.379*** | (0.13) | 0.378*** | (0.13) | 0.352*** | (0.13) | 0.359*** | (0.13) | 0.359*** | (0.13) | 0.359*** | (0.13) | 0.373*** | (0.11) |
| Network size (log) | | | | | 0.071 | (0.10) | 0.072 | (0.10) | 0.093 | (0.10) | 0.092 | (0.10) | 0.092 | (0.10) | 0.091 | (0.10) | 0.081 | (0.09) |
| **Geography** | | | | | | | | | | | | | | | | | | |
| Elevation | | | | | | | 0.001* | (0.00) | 0.001* | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | | |
| Difference between min and max of elv. | | | | | | | -0.000 | (0.00) | -0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | | |
| Distance to nearest city (in 2000) | | | | | | | -0.000 | (0.00) | -0.001 | (0.00) | -0.001 | (0.00) | -0.001 | (0.00) | -0.001 | (0.00) | | |
| Night light from 1996 to 2000 | | | | | | | -0.052** | (0.03) | -0.019 | (0.03) | -0.022 | (0.03) | -0.027 | (0.03) | -0.026 | (0.03) | | |
| **Occupation distribution in the commune** | | | | | | | | | | | | | | | | | | |
| Share of people having the same occup. | | | | | | | | | 1.206*** | (0.16) | 1.198*** | (0.16) | 1.194*** | (0.16) | 1.198*** | (0.16) | 1.121*** | (0.15) |
| Occupation diversity index | | | | | | | | | -0.699** | (0.31) | -0.726** | (0.30) | -0.779** | (0.31) | -0.718** | (0.30) | | |
| **Ethnicity** | | | | | | | | | | | | | | | | | | |
| Kinh | -0.855*** | (0.21) | -0.795*** | (0.21) | -0.825*** | (0.21) | -0.782*** | (0.22) | -0.707*** | (0.21) | -0.943*** | (0.21) | -1.505*** | (0.35) | -0.919*** | (0.22) | -0.626*** | (0.22) |
| Kinh x share | | | | | | | | | | | | | 1.093** | (0.55) | | | | |
| Share of co-ethnic in commune | | | | | | | | | | | 0.887*** | (0.24) | 0.428 | (0.35) | 0.698** | (0.28) | 0.867*** | (0.32) |
| Ethnic polarization index | | | | | | | | | | | | | | | -0.397 | (0.26) | | |
| Constant | 1.367** | (0.57) | 1.937*** | (0.58) | 1.837*** | (0.59) | 1.994*** | (0.66) | 2.819*** | (0.71) | 2.621*** | (0.73) | 2.742*** | (0.73) | 2.994*** | (0.78) | 2.292** | (1.12) |
| Observations | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | | 10487 | |
| Pseudo $R^2$ | 0.0898 | | 0.0914 | | 0.0939 | | 0.0945 | | 0.0990 | | 0.0998 | | 0.100 | | 0.1000 | | 0.125 | |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects, the last column includes commune fixed effects.

Source: Authors' calculations based on the 2008-2016 VARHS

## 6 Conclusion

In this paper, we have examined the differences in the risk sharing networks of Vietnam's ethnic minorities and majority, as a way to approach the question of the persistence of the ethnic gap in Vietnam. Exploiting a very rich rural household data set collected between 2008 and 2016 in several provinces of Vietnam, the paper first highlights important dissimilarities between Kinh and non-Kinh households in the size of these networks, as well as in their composition. The networks of the first group are on average larger, ethnically more homogenous, but more diverse in terms of occupation and geographic location than those of other ethnic groups. Thus households from the Kinh majority appear better equipped than ethnic minorities to face covariant shocks. This result is all the more important where vulnerability test shows that rural households in Vietnam are exposed to a wide range of covariant shocks, and that ethnic minorities are particularly vulnerable to these shocks.

We go further in analysing these network differences by exploring various mechanisms likely to drive these gaps. We combine survey data with census and geo-referenced data that provide detailed contextual information on the geography and the demography of sampled communes. We focus on mechanisms that are directly or indirectly linked to ethnicity and are likely to explain observed differences. Among the indirect factors, we consider household endowments, wealth, and geography; while direct mechanisms refer to effects of demography (or baseline homophily), and preference effects (or inbreeding homophily). Multivariate regressions explaining different features of risk sharing networks are estimated to disentangle these effects.

First, we find evidence that the gaps between Kinh and non-Kinh risk sharing networks are at least partly driven by the social distance that exists between ethnic groups in Vietnam, particularly between the Kinh and the non-Kinh. Indeed, households from the first group are found to have higher preferences for inbreeding homophily, i.e. for networks only composed of co-ethnics, compared to other groups, even in mixed communes. By forming such networks, they exclude ethnic minorities from their networks which are larger, more diverse geographically, and in terms of occupation. We also find evidence that the greater occupational diversity observed in Kinh networks is a consequence of this inbreeding behaviour, as Kinh occupy more diverse positions than the non Kinh in all types of communes. Results indicate furthermore that the Kinh have higher preferences for larger networks than other groups.

Along with these results that relate to household preferences, our findings indicate that the differences in the social networks of the two groups also hinge on the demographic differences between ethnic groups, and on the fact that the Kinh are by far more numerous than other groups. Consequently, they have higher opportunities to link up with other co-ethnics. This demographic effect however does not explain why Kinh networks are larger, and geographically more widely spread.

Differences in these two features (size and geographic dispersion) between Kinh and non Kinh are mostly explained by indirect effects, as the gap is absorbed by differences in other variables, such as education, wealth or by geographic conditions. Finally, in the case of network size, we find that some unobserved commune features, which are possibly related to some institutional characteristics (social cohesion for instance), are both correlated with households network size and ethnicity and partly explain the gap.

These findings must be added to those of the literature exploring the causes of Vietnam's persistent ethnic gaps. Most of these papers highlight the existence of inequalities in the endowments between ethnic groups with a strong emphasis on geographical factor, but also show differences in the returns to these endowments between Kinh and non Kinh. According to the literature, the ethnic gap will not be solved through redistributive policies alone and the inequalities need to be explored more profoundly.

Our contribution to this literature is first to show that the differences between Kinh and non Kinh risk sharing networks could be an important source of inequality in rural areas where risk is pervasive. Our paper is in line with and goes beyond the suggested evidences of ethnic segmentation in social networks (Singhal and Beck, 2015). Indeed, the lower efficiency of ethnic minorities' risk sharing networks for dealing with covariant shocks, combined with their greater vulnerability, is likely to explain why their returns to endowments are lower. Finally, we find that differences in networks cannot be explained by differences in endowments, geography or demography alone, but that they are partly rooted in the cultural and social distances that exist between the Kinh and other ethnic groups. This suggests that, unless those distances are reduced, ethnic inequalities in Vietnam are likely to persist.

**References**


Aida. 2015. 'Spatial vs. Social Network Effects in Risk Sharing,' JICA-RI Working Paper (89). Tokyo: JICA Research Institute.

Alesina, Alberto and Eliana La Ferra. 2005. 'Ethnic Diversity and Economic Performance.' *Journal of Economic Literature* 43(3): 762-800.

Arcand, Jean-Louis and Dany Jaimovich. 2014. 'Does ethnic diversity decrease economic interactions? Evidence from exchange networks in rural Gambia.' MPRA Paper 60497.

Attanasio, Orazio, Abigail Barr, Juan Camilo Cardenas,Garance Genicot and Costas Meghir. 2012. 'Risk Pooling, Risk Preferences, and Social Networks.' *American Economic Joural: Applied Economics* 4(2): 134–67.

Banerjee, Abhijit V., and Esther Duflo. 2005. 'Chapter 7 Growth Theory through the Lens of Development Economics.' *Handbook of Economic Growth* 1 (SUPPL. PART A): 473–552.

Baulch, B., T.T.K. Chuyen, D. Haughton, and J. Haughton. 2007. 'Ethnic Minority Development in Vietnam'. *Journal of Development Studies* 43(7): 1151–76.

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.

Bidart, Claire, and Johanne Charbonneau. 2012. 'How to Generate Personal Networks: Issues and Tolls for a Sociological Perspective.' HAL Working Paper 00512180. France: HAL Open Archive.

Bramoullé, Yann, and Kranton, Rachel. 2007. 'Public goods in networks.' *Journal of Economic theory* 135(1): 478-94.

Brandt, K. and Tarp, F. 2017. 'Characteristics of the VARHS Data and Other Data Sources', in Tarp, F. (ed). *Growth, Structural Transformation and Rural Change in Vietnam: A Rising Dragon on the Move*. Oxford University Press.

Burt, Ronald S. 1986. 'A note on sociometric order in the general social survey network data.' *Social Network* 8(2): 149–89.

Burt, Ronald S. (ed.) 1995. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Havard University Press. 324 p.

Carter, Michael R., Peter D. Little, Tewodaj Mogues, and Workneh Negatu. 2007. 'Poverty Traps and Natural Disasters in Ethiopia and Honduras.' *World Development* 35 (5): 835–56.

Chandrasekhar, Arun G., Cynthia Kinnan and Horacio Larreguy. 2014. 'Social networks as contract enforcement: Evidence from a lab experiment in the field.' NBER Working Paper Series 20259. Cambridge, MA: National Bureau of Economic Research.

Chantarat, Sommarat, and Christopher B. Barrett. 2012. 'Social Network Capital, Economic Mobility and Poverty Traps.' *Journal of Economic Inequality*. Vol. 10.

Coate, Stephen, and Martin Ravallion. 1993. 'Reciprocity without Commitment. Characterization and Performance of Informal Insurance Arrangements.' *Journal of Development Economics* 40 (1): 1–24.

Coleman. 1990. 'Social Capital in the Creation of Human Capital.' *American Journal of Sociology* 94 (Supplement): 59-120.

Comola, Margherita. 2010. 'The network structure of mutual support links: Evidence from rural Tanzania.' PSE Working Paper 2008-74. Paris: Paris School of Economics

Currarini, Sergio, Matthew O. Jackson and Paolo Pin. 2009. 'An Economic model of Friendship: Homophily, Minorities, and Segregation.' *Econometrica* 77(4):1003 –45.

De Weerdt, Joachim, and Stefan Dercon. 2006. 'Risk-Sharing Networks and Insurance against Illness.' *Journal of Development Economics* 81 (2): 337–56.

Dercon, Stefan. 2002. 'Income Risk, Coping Strategies and Safety Nets.' WIDER Working Paper 2002/22. Helsinki: UNU-WIDER.

Déry, Steve. 2003. 'Distinctions ethniques et déforestation au Việt Nam', *Etudes rurales* 165-6: 81-102.

Dương Bích Hạnh. 2008. 'Contesting Marginality: Consumption, Networks, and Everyday Practice among Hmong Girls in Sa Pa, Northwestern Vietnam.' *Journal of Vietnamese Studies* 3:231-260.

Epprecht, Michael, Daniel Muller, and Nicholas Minot. 2009. 'How remote are Vietnam's ethnic minorities? An analysis of spatial patterns of poverty and inequality.'. *The Annals of Regional Science* 46(2): 349-68.

Esteban, J., and Ray, D. 2011. 'Linking Conflict to Inequality and Polarization.' *American Economic Review* 101(4): 1345-74.

Fafchamps, Marcel. (ed.) 2003. *Rural Poverty, Risk and Development*. Cheltenham: Edward Elgar Publishing.

———. 2011. *Risk Sharing between Households. Handbook of Social Economics* 1 (1 B): 1255–79.

Fafchamps, Marcel, and Flore Gubert. 2007. 'The Formation of Risk Sharing Networks.' *Journal of Development Economics* 83 (2): 326–50.

Fafchamps, Marcel, Christopher Udry, and Katherine Czukas. 1998. 'Drought and Saving in West Africa: Are Livestock a Buffer Stock?' *Journal of Development Economics* 55 (2): 273–305.

Fischer, Isabel, Tina Beuchelt, Tom Dufhues, and Gertrud Buchenrieder. 2010. 'Risk mamagement networks of ethnic minorities in Vietnam.' *Asia-Pacific Development Journal* 17(2): 93-118.

Fitzsimons, Emla, Bansi Malde, Marcos Vera-Hernández. 2018. 'Group Size and the Efficiency of Informal Risk Sharing.' The Economic Journal 128(612): 575-608.

Genicot, Garance, and Debraj Ray. 2003. 'Group formation in risk-sharing arrangements', *Review of Economic Studies* 70(1): 87-113.

Goldstein, Markus, Alain De Janvry, and Elisabeth Sadoulet. 2005. 'Is a Friend in Need a Friend Indeed? Inclusion and Exclusion in Mutual Insurance Networks in Southern Ghana.' *Insurance Against Poverty*: 217-48.

Goodman, S., BenYishay, A., Runfola, D., 2016. Overview of the geo Framework. AidData. Available online at geo.aiddata.org.

Grimard, Franque. 1997. 'Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire.' *Journal of Development Economics* 53(1997): 391–422.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2007. 'Why Ethnic Diversity Undermine Public Goods Provision.' *American Political Science Review* 101(4).

Hardy, Andrew. 2003. Red Hills: Migration and the State in the Highlands of Vietnam. NIAS Press, Copenhagen.

Hardy, Andrew and Nguyen, Van Chinh. 2004. 'Gérer la "question ethnique" sur les hautes terres : un exercice d'équilibriste'. In S. Dovert, and B. De Treglodé (eds.), *Viêt Nam contemporain*. Bangkok: IRASEC/Les Indes savantes : 383-432.

Hasegawa, Masako. 2010. 'Risk Coping Measures against Different Types of Shocks: Empirical Evidence from Vietnam Household Living Standard Survey.' OSIPP Discussion Paper. Osaka: Osaka School of International Public Policy, Osaka University.

Henderson, J. V., Storeygard, A., and Weil, D. 2012. 'Measuring economic growth from outer space. *American Economic Review*, 102 (2), 994–1028.

Hodler, R., and Raschky, P. 2014b. 'Regional favoritism'. *The Quarterly Journal of Economics*, 129 (2), 995–1033.

Horowitz, Donald L. (ed.) 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Imai, K. S., Gaiha, R., and Kang, W. 2011. 'Poverty, inequality and ethnic minorities in Vietnam.' *International Review of Applied Economics*, 25, 249–282

Institute for Studies of Society, Economics and Environment (ISEE). 2011. 'Portrayal of Ethnic Minorities.' Printed Newspapers. Hanoi: The Gioi Publishers

Jackson, Matthew O., Brian Rogers and Yves Zenou. 2016. 'The Economic Consequences of Social Network Structure.' IFN Working Paper No.116. Stockholm: Institute of Industrial Economics.

Jaramillo, Fernando, Hubert Kempt and Fabien Moizeau. 2015. 'Heterogeneity and the formation of risk-sharing coalitions.' *Journal of Development Economics* 114: 79-96.

Johny, Judit, Bruno Wichmann and Brent M. Swallow. 2017. 'Characterizing social networks and their effects on income diversification in rural Kerala, India.' *World Development* 94: 375-92.

Karlan, Dean, Robert Osei-Akoto, and Christopher Udry. 2014. 'Agricultural Decision after Relaxing Credit and Risk Constraints.' *The Quarterly Journal of Economics* 2 (129): 597–652.

Khwaja, Asim. 2009. 'Can good projects succeed in bad communities.' *Journal of Public Economics* 93(7-8): 899-916.

Ligon, E. and Schechter, L. 2003. 'Measuring vulnerability.' *Economic Journal*, 113(486): 95-102.

McElwee, Pamela. 2004. 'Becoming Socialist or Becoming Kinh? Government Policies for Ethnic Minorities in the Socialist Republic of Vietnam'. In C.R. Duncan (ed.), *Civilizing the Margins : Southeast Asian Government Policies for the Development of Minorities*. Ithaca: Cornell University Press.

McElwee, Pamela. 2008. '"Blood Relatives" or Uneasy Neighbors? Kinh Migrant and Ethnic Minority Interactions in the Trường Sơn Mountains.' *Journal of Vietnamese Studies* 3:81-116.

McPherson, Smith-Loven and Cook. 2001. 'Birds of a Feather: Homophily in Social Networks.' *Annual Review of Sociology*, Vol. 27:415-44.

Michaud, Jean. 2008. 'Flexibilité de l'économie chez les Hmong de la haute région du Viêt-nam septentrional'. *Aséanie* 27:151-83.

Miguel, Edward and Mary Kay Gugerty. 2005. 'Ethnic diversity, social sanctions, and public goods in Kenya.' *Journal of Public Economics* 89: 2325–68.

Montalvo, José, G., and Marta Reynal-Querol. 2005. 'Ethnic Polarization, Potential Conflict, and Civil Wars.' *American Economic Review* 95(3): 796-816.

Morduch, Jonathan. 1994. 'Poverty and vulnerability', *American Economic Review*, 84(2): 221-25

Morduch, Jonathan. 1995. 'Income Smoothing and Consumption Smoothing.' *Journal of Economic Perspectives*, 9(3): 103-14.

Nguyen Hoa Thi Minh, Tom Kompas, Trevor Breusch, and Michael B. Ward. 2017. 'Language, Mixed Communes, and Infrastructur: Sources of Inequality and Ethnic Minorities in Vietnam.' *World Development*, 96: 145-162.

Pasquier-Doumer, Laure and Fiorella Risso Brandon. 2015. 'Aspiration Failure: A Poverty Trap for Indigenous Children in Peru?.' *World Development* 72: 208–23.

Rockenbauch, Till and Patrick Sakdapolrak. 2017. 'Social networks and the resilience of rural communities in the Global South: a critical review and conceptual reflection.' *Ecology and Society* 22(1): 10.

Saint-Macary, Camille 2014. 'Microeconomic Impacts of Institutional Change in Vietnam's Northern Uplands: Empirical Studies on Social Capital, Land and Credit Institutions.' *Development Economics and Policy*, vol. 71, Peter Lang Verlag, Frankfurt am Main, Germany.

Singhal, Saurabh, and Ulrik Beck. 2015. 'Ethnic Disadvantage In Vietnam: Evidence Using Panel Data'. WIDER Working Paper 2015/097. Helsinki: UNU-WIDER.

Townsend, Rober. 1994. 'Risk and Insurance in Village India.' *Econometrica,* 62(3): 539-91.

Tran, Van Q. 2015. 'Household's Coping Strategies and Recoveries from Shocks in Vietnam.' *Quarterly Review of Economics and Finance* 56. Board of Trustees of the University of Illinois: 15–29.

United Nations Development Programme. 1998. 'The Dynamics of Internal Migration in Vietnam.' Hanoi: UNDP Discussion Paper 1.

Van de Walle, Dominique, and Dileni Gunewardena. 2001. 'Souces of Ethnic Inequality in Vietnam.' *Journal of Development Economics* 65 (1): 177–207.

Vasavakul, T. 2003. 'Language policy and ethnic relations in Vietnam'. In E. Brown and S. Ganguly (Eds.), *Fighting Words: Language Policy and Ethnic Relations in Asia*. Cambridge: MIT Press.

Vu Hong Phong. 2013. 'A mixed methods analysis of drinking cultures in Northern Vietnam'. PhD thesis. Department of Sociology, University of Essex.

World Bank. 2012. 'Well Begun, Not Yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges.' 2012 Vietnam Poverty Assessmentc Report, no. 74910: 177. Washington, DC: World Bank.

Zylberberg, Yanos. 2010. 'Do tropical typhoons smash community ties? Theory and Evidence from Vietnam.' PSE Working paper 2010/35. Paris: Paris School of Economics.

# Appendix

# Appendix 1

**Table A. Descriptive statistics of households**

| Variable | Kinh (n=8858) | Non-Kinh (n=2317) | Difference (T-test) | Kinh (n=5819) | Non-Kinh (=2159) | Difference (T-test) |
|---|---|---|---|---|---|---|
| | In all 428 communes | | | In 306 mixed communes with Kinh | | |
| Age of head | 54.24 | 47.48 | 6.76*** | 53.94 | 47.50 | 6.44*** |
| Male headed | 0.75 | 0.89 | -0.14*** | 0.76 | 0.89 | -0.14*** |
| Household size | 3.99 | 5.15 | -1.16*** | 3.96 | 5.12 | -1.17*** |
| Head or spouse born in the village | 0.83 | 0.87 | -0.05*** | 0.79 | 0.87 | -0.08*** |
| Farming (agriculture. forestry and aquaculture) | 0.63 | 0.83 | -0.19*** | 0.63 | 0.82 | -0.19*** |
| Cannot read/write or never go to school | 0.06 | 0.30 | -0.23*** | 0.05 | 0.29 | -0.24*** |
| Completed lower primary | 0.21 | 0.26 | -0.05*** | 0.20 | 0.26 | -0.06*** |
| Completed lower secondary | 0.47 | 0.34 | 0.13*** | 0.48 | 0.35 | 0.13*** |
| Completed upper secondary | 0.25 | 0.10 | 0.15*** | 0.27 | 0.10 | 0.17*** |
| Log of durable assets | 9.61 | 8.87 | 0.74*** | 9.70 | 8.90 | 0.81*** |
| Share of co-ethnics in the commune | 0.95 | 0.52 | 0.43*** | 0.92 | 0.50 | 0.42*** |
| Share of people having the same occupation in the commune | 0.35 | 0.66 | -0.31*** | 0.35 | 0.65 | 0.30*** |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

**Table B. Definition of shocks**

| Variable | Definition |
|---|---|
| Covariant shock | A dummy variable indicating if the household has suffered a least one spatially covariant natural or economic shock. |
| Natural shock | A dummy indicating if the household was exposed to any of the following events: floods, landslides, typhoons, storms, droughts, pest infestation, crop disease and avian flu |
| Economic shock | A dummy indicating if the household was exposed to any of the following events: changes in crop prices, key input prices, prices of food or other essential commodities |
| Idiosyncratic shock | A dummy indicating if the household suffered any of the following events: illness, injury or death of a family member; unemployment, unsuccessful investment; loss of land; crime/robbery/theft; divorce, family disputes. |

Source: VARHS questionnaire.

## Appendix 2 Vulnerability estimation

There are different ways of measuring vulnerability, depending on how vulnerability is defined: vulnerability as expected poverty (VEP), vulnerability as expected utility (VEU), and vulnerability as exposure to risk (VER). A rich discussion has emphasized the advantages and disadvantages of each method (Hoddinott and Quisumbing, 2003; Ligon and Schechter, 2004; Gaiha and Imai, 2008). We use the concept of vulnerability as low expected utility proposed by Ligon and Schechter (2003) in which, vulnerability is decomposed into different types of risks which suits our purpose of this study.

Vulnerability is defined as the difference between household's utility obtained from some certainty equivalent consumption and the expected utility of consumption.

$$V^i(c) = U^i(z) - EU^i(c^i)$$

The certainty equivalent consumption (z) is similar to the choice of a "poverty line", such that if the consumption expenditure of a household is lower than this level, the household is considered vulnerable and vice versa. The vulnerability is then decomposed into two terms:

$$V^i(c) = \left[U^i(z) - U^i(Ec^i)\right] + \left[U^i(Ec^i) - EU^i(c^i)\right]$$

The first term measures poverty, which is the difference between utility of certainty equivalent consumption and utility from expected consumption ($Ec^i$) of household i. The second term measures the risk faced by household i; which is further decomposed into aggregate risk and idiosyncratic risks as follows:

$$\begin{aligned} V^i(c) = {} & \left[U^i(z) - U^i(Ec^i)\right] && \text{(Poverty)} \\ & + \left[U^i(Ec^i) - EU^i\left(E(c^i|\bar{x})\right)\right] && \text{(Aggregate risk)} \\ & + \left[EU^i\left(E(c^i|\bar{x})\right) - EU^i(c^i)\right] && \text{(Idiosyncratic risk)} \end{aligned}$$

Of which, $E(c^i|\bar{x})$ is the expected value of consumption, conditional on a vector of aggregate variables $\bar{x}$. To estimate risk, Ligon and Schechter (2003) use the variation over time, thus, t is introduced into the equation such that $c_t^i$ is the consumption of household i at time t; $x_t^i$ is a vector of household i's idiosyncratic variables and $\bar{x}_t$ is a vector of aggregate variables. Accordingly, the vulnerability function36 is:

$$\begin{aligned} V^i(c) = {} & \left[U^i(Ec) - EU^i(Ec_t^i)\right] && \text{(Poverty)} \\ & + \left[U^i(Ec_t^i) - EU^i\left(E(c_t^i|\bar{x}_t)\right)\right] && \text{(Aggregate risk)} \\ & + \left[EU^i\left(E(c_t^i|\bar{x}_t)\right) - EU^i\left(E(c_t^i|\bar{x}_t, x_t^i)\right)\right] && \text{(Idiosyncratic risk)} \\ & + \left[EU^i\left(E(c_t^i|\bar{x}_t, x_t^i)\right) - EU^i(c_t^i)\right] && \text{(Unexplained risk \& measurement error)} \end{aligned}$$

[36] See Ligon and Schechter (2003) for a detailed description of vulnerability estimation

It is noted that the utility function applied in Ligon and Schechter's has the form of $U^i(c) = (c^{1-\gamma})/(1-\gamma)$. Of which, $\gamma$ is the household's relative risk aversion and it is chosen at 2 in estimation, based on the empirical studies. They also normalize c, thus, the average of consumption for all households across all periods (z) equals 1. The parameters of $E(c_t^i|\bar{x}_t, x_t^i)$ can be estimated by linear equation:

$$E(c_t^i|\bar{x}_t, x_t^i) = \alpha_i + \eta_t + \beta x_t^i$$

In which, $\alpha_i$ is the time invariant household's characteristics; $\eta_t$captures the time effects across all households; and $\beta$ is the effect of household's characteristics and/or other observable factors on consumption.

We estimate the vulnerability using food consumption expenditures because the data does not include non-food consumption. Then, vulnerability is decomposed into poverty, aggregate risk, idiosyncratic risk and unexplained risk. Within the idiosyncratic risk; we consider the variance in household's income and changes in number of workers in each family as the sources of idiosyncratic shocks, of which, the changes in number of worker are orthogonalized to changes in income. The confidence interval is bootstrapped for each component. Lastly, we also use the same weight scales for adult equivalent consumption as used by Townsend (1994); Ligon and Schechter (2003). In which, adult male is assigned a weight of 1 and adult female a weight of 0.9 (adult means sixteen or older). Children aged 0 to 4 count as 0.32, aged 5 to 9 as 0.52 and aged 10 to 15 as 0.67.

**Table C. Correlates and decomposition of vulnerability in food consumption**

| | Vulnerability | Poverty | Aggregate risk | Idiosyncratic risk | Unexplained risk |
|---|---|---|---|---|---|
| | 0.785*** | 0.278*** | 0.040*** | 0.005*** | 0.462*** |
| | [ 0.734, 0.85] | [0.262, 0.295] | [0.035, 0.047] | [0.004, 0.008] | [0.421, 0.516] |
| Variable | Coef | Coef | Coef | Coef | Coef |
| Age | -0.003 | 0.002 | 0.000 | 0.000 | -0.005 |
| | (0.019) | (0.008) | (0.000) | (0.000) | (0.014) |
| Age sq. | 0.000 | -0.000 | -0.000 | -0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Head is male+ | 0.060 | 0.031 | 0.001 | 0.000 | 0.028 |
| | (0.071) | (0.036) | (0.001) | (0.001) | (0.047) |
| HH size | 0.104*** | 0.074*** | 0.002*** | 0.000 | 0.027 |
| | (0.040) | (0.017) | (0.001) | (0.000) | (0.028) |
| Head completed primary school+ | -0.206 | -0.144** | -0.005** | 0.000 | -0.060 |
| | (0.187) | (0.062) | (0.002) | (0.001) | (0.151) |
| Head completed lower secondary+ | -0.577*** | -0.344*** | -0.011*** | 0.001 | -0.224* |
| | (0.175) | (0.069) | (0.002) | (0.002) | (0.13) |
| Head completed upper secondary+ | -0.891*** | -0.536*** | -0.017*** | 0.007 | -0.346** |
| | (0.192) | (0.073) | (0.003) | (0.006) | (0.147) |
| Head or Spouse are village native+ | -0.033 | 0.040 | 0.001 | -0.001 | -0.073 |
| | (0.197) | (0.061) | (0.001) | (0.001) | (0.155) |
| Head is Kinh+ | -0.754*** | -0.429*** | -0.014*** | -0.001 | -0.311** |
| | (0.260) | (0.141) | (0.004) | (-0.002) | (0.137) |
| Farming is the main source of income + | 0.223** | 0.153*** | 0.005** | 0.001 | 0.064 |
| | (0.095) | (0.049) | (0.002) | (0.003) | (0.06) |
| Irrigated land area | -0.000** | -0.000** | -0.000** | -0.000 | -0.000* |
| | (0.000) | 0.000 | (0.000) | (0.000) | (0.000) |
| #workers | -0.103** | -0.043** | -0.001** | 0.002** | -0.060** |
| | (0.044) | (0.018) | (0.001) | (0.001) | (0.032) |
| Commune dummies | Yes | Yes | Yes | Yes | Yes |
| R2 | 0.581 | 0.6795 | 0.680 | 0.081 | 0.469 |

Note: Numbers in parenthesis are bootstrapped standard errors and those in brackets are 95% confidence intervals.
Significance levels: * p<0.10, ** p<0.05, *** p<0.01.
Source: Authors' calculations based on the VARHS 2008-2016.

## Appendix 3 Censored network bias

Regarding the limitation issue of using censored network, we compare the difference between the size of the risk sharing network between the censored question (up to three most important contacts) and question without upper bound. We find that on average, people list 2 contacts when they are asked to list a maximum of 3 and 4.3 contacts if there is no limitation (with a standard deviation of 5.5). Since the number of contacts varies considerably when there is no limit, we drop the outliers which are higher than the mean by three standard deviations; those outliers account for 1.5 per cent of the observations. Without the outliers, the size of uncensored network is 3.8 with the maximum of 21. Then, we take the difference (in log) between the censored and uncensored network and run an OLS regression to examine the correlates of household's characteristics with this gap. It is seen that around 60 per cent of the observations give the same number of contacts regardless of the questions and the mean of the difference is 1.8 with a standard deviation of 3.3. Regression results show that male-headed households, households with higher educated heads, and better off households have bigger network than the censored one. The gap is also positively correlated with Kinh households. When we add controls for structure of the network in the regression, we find that households with high proportion of same co-ethnic or occupational similarity contacts seem to have bigger size than their censored answer. Thus, it could imply that if they described their network based on the uncensored one, they could have had a less homophily structure. This finding corresponds to our discussion above that we might underestimate the diversity in Kinh's network because Kinh households seem to have bigger networks than three and out of the three contacts, they might ask for help from more diverse sources than what we find in their structure limited within the most three important contacts.

**Table D. Correlates of network size bias**

| Dependent variable: log of network size difference | (1) OLS | (2) OLS | (3) OLS | (4) OLS |
|---|---|---|---|---|
| Age of HH head | -0.000 | 0.002 | 0.002 | 0.001 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| (Age of head)2 | 0.000 | -0.000 | -0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Head is male | 0.068*** | 0.071*** | 0.071*** | 0.064*** |
| | (0.019) | (0.020) | (0.020) | (0.020) |
| HH size | 0.002 | 0.005 | 0.005 | 0.003 |
| | (0.005) | (0.005) | (0.005) | (0.005) |
| Head completed primary school | 0.027 | 0.032 | 0.031 | 0.032 |
| | (0.026) | (0.028) | (0.028) | (0.028) |
| Head completed lower secondary | 0.045* | 0.042 | 0.042 | 0.045 |
| | (0.027) | (0.028) | (0.028) | (0.028) |
| Head completed upper secondary | 0.118*** | 0.128*** | 0.129*** | 0.140*** |
| | (0.031) | (0.033) | (0.033) | (0.033) |
| Head is farmer | -0.014 | -0.023 | -0.023 | -0.032 |
| | (0.019) | (0.019) | (0.019) | (0.020) |
| Head or Spouse is village native | -0.023 | -0.031 | -0.031 | -0.030 |
| | (0.030) | (0.031) | (0.031) | (0.031) |
| Kinh | 0.114** | 0.097* | 0.108** | 0.115** |
| | (0.048) | (0.052) | (0.051) | (0.051) |
| Asset value (log) | 0.032*** | 0.034*** | 0.034*** | 0.036*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Proportion of co-ethnic in the network | | 0.132*** | | |
| | | (0.051) | | |
| Proportion of within-village contacts in the network | | | 0.026 | |
| | | | (0.020) | |
| Proportion of same occupation contacts in the network | | | | 0.115*** |
| | | | | (0.019) |
| Constant | -0.272* | -0.442** | -0.337** | -0.389** |
| | (0.152) | (0.175) | (0.166) | (0.167) |
| Commune FE | Yes | Yes | Yes | Yes |
| Year dummy | Yes | Yes | Yes | Yes |
| Observation | 1112 | 10453 | 10453 | 10453 |
| R-squared | 0.137 | 0.142 | 0.142 | 0.145 |

Note: Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Authors' calculations based on the 2008-2016 VARHS.

## Appendix 4 Table E. Determinants of risk sharing network's size and similarity by ethnicity disaggregation

| Depedent variable | Network size | | | | Ethnic similarity | | | | Geographical proximity | | | | Occupational similarity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS (1) | | OLS (2) | | Tobit (3) | | Tobit (4) | | Tobit (5) | | Tobit (6) | | Tobit (7) | | Tobit (8) | |
| Age of HH head | -0.006** | (0.00) | -0.006** | (0.00) | -0.034 | (0.04) | -0.036 | (0.04) | -0.007 | (0.01) | -0.008 | (0.01) | 0.019 | (0.02) | 0.018 | (0.02) |
| (Age of head)2 | 0.000** | (0.00) | 0.000** | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) | -0.000 | (0.00) | 0.000 | (0.00) | -0.000** | (0.00) | -0.000** | (0.00) |
| Head is male | 0.049*** | (0.02) | 0.048*** | (0.02) | 0.105 | (0.22) | 0.108 | (0.22) | 0.154** | (0.07) | 0.156** | (0.07) | 0.402*** | (0.09) | 0.404*** | (0.09) |
| Head completed primary school | 0.035 | (0.02) | 0.035 | (0.02) | -0.090 | (0.31) | -0.039 | (0.31) | -0.084 | (0.10) | -0.093 | (0.10) | -0.009 | (0.15) | -0.005 | (0.15) |
| Head completed lower secondary | 0.057** | (0.02) | 0.057** | (0.02) | -0.143 | (0.32) | -0.075 | (0.32) | -0.214* | (0.11) | -0.216* | (0.11) | -0.232 | (0.15) | -0.223 | (0.15) |
| Head completed upper secondary | 0.099*** | (0.03) | 0.099*** | (0.03) | -0.115 | (0.38) | -0.040 | (0.37) | -0.501*** | (0.13) | -0.508*** | (0.18) | -0.872*** | (0.16) | -0.867*** | (0.16) |
| Head is farmer | -0.005 | (0.02) | -0.004 | (0.02) | 0.270 | (0.17) | 0.267 | (0.17) | 0.165*** | (0.06) | 0.166*** | (0.06) | 0.008 | (0.12) | 0.014 | (0.12) |
| Head or Spouse is village native | 0.017 | (0.03) | 0.016 | (0.03) | -0.084 | (0.30) | -0.073 | (0.30) | 0.173* | (0.09) | 0.176* | (0.09) | -0.052 | (0.13) | -0.058 | (0.13) |
| HH size | -0.004 | (0.00) | -0.004 | (0.00) | -0.097** | (0.05) | -0.104** | (0.05) | 0.031* | (0.02) | 0.029* | (0.08) | 0.151*** | (0.02) | 0.150*** | (0.02) |
| Asset value (log) | 0.029*** | (0.00) | 0.029*** | (0.00) | 0.072* | (0.04) | 0.074* | (0.09) | -0.091*** | (0.02) | -0.091*** | (0.02) | -0.111*** | (0.02) | -0.110*** | (0.02) |
| Censor | | | | | 0.525* | (0.31) | 0.512* | (0.31) | -0.037 | (0.09) | -0.039 | (0.09) | 0.358*** | (0.13) | 0.353*** | (0.13) |
| Network size (log) | | | | | -0.109 | (0.23) | -0.085 | (0.29) | 0.047 | (0.07) | 0.048 | (0.07) | 0.091 | (0.10) | 0.095 | (0.10) |
| Elevation | 0.000 | (0.00) | 0.000 | (0.00) | 0.001 | (0.00) | 0.000 | (0.00) | 0.001* | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) |
| Difference between min and max of elv. | -0.000 | (0.00) | -0.000 | (0.00) | 0.001 | (0.00) | 0.001 | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | 0.000 | (0.00) | 0.000 | (0.00) |
| Distance to nearest city (in 2000) | -0.000* | (0.00) | -0.000** | (0.00) | -0.005*** | (0.00) | -0.005*** | (0.00) | -0.000 | (0.00) | -0.000 | (0.00) | -0.001 | (0.00) | -0.001 | (0.00) |
| Night light from 1996 to 2000 | -0.004 | (0.05) | -0.004 | (0.01) | 0.045 | (0.06) | 0.022 | (0.06) | -0.025* | (0.02) | -0.026* | (0.02) | -0.027 | (0.03) | -0.029 | (0.03) |
| Number of villages | | | | | | | | | 0.002 | (0.00) | 0.003 | (0.00) | | | | |
| Share of people having the same occupation | | | | | | | | | | | | | 1.206*** | (0.16) | 1.195*** | (0.16) |
| Occupation diversity index | | | | | | | | | | | | | -0.711** | (0.301) | -0.733** | (0.30) |
| Share of co-ethnic in commune | -0.045 | (0.05) | -0.113 | (0.10) | 2.460*** | (0.51) | 4.295*** | (0.97) | 0.070 | (0.19) | 0.160 | (0.42) | 0.788*** | (0.295) | 1.210** | (0.50) |
| Ethnic polarization index | -0.047 | (0.05) | -0.069 | (0.05) | -1.188** | (0.49) | -0.726 | (0.52) | 0.251 | (0.22) | 0.248 | (0.24) | -0.365 | (0.265) | -0.251 | (0.29) |
| 2.Thai | -0.034 | (0.05) | -0.083 | (0.09) | -1.282* | (0.68) | 0.499 | (1.14) | 0.047 | (0.28) | 0.926* | (0.50) | 0.832*** | (0.312) | 1.723*** | (0.48) |
| 3.Tay | 0.050 | (0.07) | -0.012 | (0.10) | -1.737*** | (0.66) | -0.888 | (0.93) | 0.394* | (0.21) | 0.194 | (0.44) | 1.469*** | (0.392) | 1.394** | (0.56) |
| 4.Hmong | -0.089 | (0.06) | -0.013 | (0.11) | -1.208 | (0.79) | -2.204* | (1.17) | 0.259 | (0.36) | -0.891 | (0.59) | 0.692* | (0.384) | 0.429 | (0.60) |
| 5.Others | -0.076* | (0.04) | -0.140* | (0.08) | -0.530 | (0.49) | 1.137 | (0.78) | 0.196 | (0.17) | 0.334 | (0.36) | 0.849*** | (0.245) | 1.233*** | (0.43) |
| 2.Thai x share | | | 0.088 | (0.14) | | | -3.322** | (1.63) | | | -1.318** | (0.65) | | | -1.509** | (0.70) |
| 3.Tay x share | | | 0.189 | (0.2) | | | -0.458 | (1.85) | | | 1.194 | (1.36) | | | 1.787 | (1.39) |
| 4.Hmong x share | | | -0.104 | (0.15) | | | 2.349 | (1.63) | | | 1.836** | (0.78) | | | 0.308 | (0.98) |
| 5.Others x share | | | 0.130 | (0.14) | | | -3.547*** | (1.34) | | | -0.305 | (0.53) | | | -0.746 | (0.72) |
| Constant | 0.914*** | (0.135) | 0.975*** | (0.16) | 8.454*** | (1.98) | 7.016*** | (2.05) | 3.299*** | (0.54) | 3.259*** | (0.66) | 1.778** | (0.766) | 1.389 | (0.85) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observations | 11175 | 11175 | 10487 | 10487 | 10487 | 10487 | 10487 | 10487 |
| R2/ Pseudo $R^2$ | 0.108 | 0.108 | 0.174 | 0.178 | 0.0436 | 0.0445 | 0.100 | 0.100 |

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All equation includes time and district fixed effects

Source: Authors' calculations based on the 2008-2016 VARH

## Résumé

Au cours des trente dernières années, le Vietnam a enregistré un succès remarquable en termes de réduction de la pauvreté et de croissance économique. Toutefois, la réduction des inégalités horizontales, et en particulier les inégalités entre groupes ethniques, demeure un défi pour ce pays. Cette thèse de doctorat porte sur les inégalités horizontales et la pauvreté des minorités ethniques au Vietnam en privilégiant trois axes thématiques : les effets du principe de participation sur des programmes de réduction de la pauvreté ; les écarts de revenus du travail selon le sexe ou l'appartenance ethnique, et enfin, le lien entre ethnicité et formation des réseaux sociaux. Le premier chapitre porte sur la participation. Ce principe est-il effectivement appliqué dans les programmes de réduction de la pauvreté ciblés sur les zones montagneuses et les minorités ethniques ? La participation a-t-elle un effet sur les résultats économiques et sociaux des ménages ? Le deuxième chapitre examine les écarts de rémunérations entre hommes et femmes et entre groupes ethniques sur le marché du travail vietnamien. Dans quelle mesure ces écarts peuvent être expliqués par des inégalités de dotations (éducation, capital, etc.) ou de rendement de ces dotations. Enfin, le troisième chapitre explore les différences entre minorités et majorité ethniques quant aux réseaux de partage des risques, Il et met en lumière l'importance de la distance culturelle et sociale dans la formation des réseaux sociaux. Ces trois chapitres démontrent la persistance d'inégalités horizontales au Vietnam. Ils laissent augurer que ces inégalités persisteront tant que l'accès aux ressources physiques ne sera pas comblé et que la distance culturelle et sociale n'aura pas été résorbée.



## Mots Clés

ethnicité, inégalité, marché du travail, réseau social, Viet Nam



## Abstract

Vietnam has achieved remarkable success in poverty reduction and economic growth over the last three decades. However, horizontal inequality in general and ethnic inequality in particular remains a persistent challenge for Vietnam. This PhD thesis studies horizontal inequalities and ethnic poverty in Vietnam along three axes: effects of community participation in poverty reduction programs, earnings gaps by gender and ethnicity, and the formation of social networks by ethnic affiliation. The first chapter inspects if the motto of community participation approach takes place effectively in poverty reduction programs targeted at mountainous areas and ethnic minorities. It evaluates the effect of community participation on households' economic and welfare outcomes. The second chapter examines whether gender and ethnic earnings gaps exist on the Vietnamese labour market and how much of these gaps can be explained by differences in endowments and returns to endowments. Lastly, the third chapter explores the disparities in risk sharing networks between the ethnic minority and majority. It highlights the importance of cultural and social distance in social network formation. Taken together, the three chapters of this thesis provide evidence of horizontal inequality existence in Vietnam. The thesis suggests that these horizontal inequality will be likely to persist if differences in physical endowments and socio-cultural distance are not reduced.



## Keywords

ethnicity, inequality, labor market, social network, Viet Nam