



WIDER Working Paper 2018/134

Ethnicity and risk sharing network formation

Evidence from rural Vietnam

Quynh Hoang,¹ Laure Pasquier-Doumer,² and Camille Saint-Macary²

October 2018

Abstract: Ethnic inequality remains a persistent challenge for Vietnam. This paper aims at better understanding this ethnic gap through exploring the formation of risk sharing networks in rural areas. It first investigates the differences in risk sharing networks between the ethnic minorities and the Kinh majority, in terms of size and similarity attributes of the networks. Second, it relies on the concept of ethnic homophily in link formation to explain the mechanisms leading to those differences. In particular, it disentangles the effect of demographic and local distribution of ethnic groups on risk-sharing network formation from cultural and social distance between ethnic groups, while controlling for the disparities in the geographical environment. Results show that ethnic minorities have smaller and less diversified networks than the majority. This is partly explained by differences in wealth and in the geographical environment. But ethnicity also plays a direct role in risk-sharing network formation through the combination of preferences to form a link with people from the same ethnic group (inbreeding homophily) and the relative size of ethnic groups conditioning the opportunities to form a link (baseline homophily). 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. This evidence suggests that inequalities among ethnic groups in Vietnam 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) and the 6th Atelier DIAL-GREThA in Paris (June 2018), for the insightful comments and questions provided.

¹ PSL, Université Paris-Dauphine, Paris, France, UMR 225 DIAL, Institut de Recherche pour le Développement (IRD), Paris, France, corresponding author, email: quynhhoang176@gmail.com; ² UMR 225 DIAL, Institut de Recherche pour le Développement (IRD), Paris, France, Université Paris-Dauphine, Paris, France.

This study has been prepared within the UNU-WIDER project on ‘[Structural transformation and inclusive growth in Vietnam](#)’.

Copyright © UNU-WIDER 2018

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-576-3 <https://doi.org/10.35188/UNU-WIDER/2018/576-3>

Typescript prepared by Ans Vehmaanperä.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

Accounting for only 15 per cent of the population but 70 per cent of Viet Nam'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.¹ 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 Viet Nam 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 gap 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 Viet Nam'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 (Rockenbach and Sakdapolrak, 2017).

In addition to other factors, ethnicity is potentially an important factor affecting how a social network is formed (Grimard, 1997). This could lead to dissimilarities in the social networks between ethnic minority and majority groups in Viet Nam, and affects how ethnic minority and majority groups are able to cope with shocks. Due to their generally lower social positions, stronger isolation, and greater social distance with the Kinh majority, ethnic minorities constitute social networks which may strongly differ in size and structure from those of the majority.

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 per cent 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

¹ The most important programmes are the Programme 135 (P135) and the Hunger Eradication and Poverty Reduction (HEPR).

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. It could lead to higher exclusion of non-Kinh people from Kinh risk sharing network than predicted by their demography and local distribution. Thus, demographic, local distribution and social distance across ethnic groups may explain differences in similarity patterns across ethnic group social networks.

In this paper, we first examine whether risk sharing networks from majority ethnic group and from minority ethnic groups differ in size and similarity patterns. We observe that ethnic minority groups are embedded in smaller risk sharing networks, which implies their 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 to cope with covariant shocks. Indeed, covariant shocks, by definition, affect people from the same geographical area and usually the farmers. Starting from these observations, this paper aims at understanding, in a second step, which mechanisms are at play with a focus on baseline and inbreeding homophily. It tries to disentangle the effect of demographic and local distribution of ethnic groups on risk sharing network formation (baseline homophily) from cultural and social distance between ethnic groups (inbreeding homophily), while controlling for the disparities in the geographical environment which can affect the cost of interacting.

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 Viet Nam. 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 Viet Nam.²

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

² 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.

analysis. We present our estimation strategy and results in section 5 and briefly conclude in Section 6.

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 Weerd 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. 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 (Khawaja, 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 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 affects 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 Viet Nam 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 Viet Nam

This question is particularly relevant in the context of Viet Nam, 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 Viet Nam. 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 Viet Nam. Until the Independence, mixed communes with Kinh and ethnic minority people were rare (Déry, 2003).³ 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

³ However, communes composed of a single ethnic group were an exception as different ethnic minority groups were living next to one another (Vasavakul, 2003).

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). 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 Viet Nam. The sociologist Vu Hong Phong (2013) conducted field work on interethnic relationship in Muong Khen, a rural town in northwest mountainous Viet Nam 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 Viet Nam 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 mồi [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: 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 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 Viet Nam, 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 Viet Nam Access to Resources Household Survey (VARHS).⁴ 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 Viet Nam 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 Viet Nam Census of 2009⁵ 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 Viet Nam'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 GeoQuery⁶ to characterize households' 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

In this paper, 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

⁴This 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).

⁵This is the latest Population and Housing Census in Viet Nam, conducted in 2009 by General Statistics Office with technical and financial support from the UNFPA. The census covered all Vietnamese citizens usually residing in Viet Nam 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.

⁶GeoQuery is an open source providing access to custom subsets of spatial data. Data requests are available at <http://geo.aiddata.org/query>.

contacts that he/she would ask for help in case of cash emergency;⁷ 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 2. 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 ties⁸; and ii) the proportion of within-village contacts in their risk sharing network.⁹ 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

⁷ 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.

⁸ 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.

⁹ 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.

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 Viet Nam, 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 Viet Nam. 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 Viet Nam, 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 intermarriage with a Kinh partner that may affect the composition of the 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.

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, Viet Nam is an ethnically diverse country (Figure 1). The Kinh live almost everywhere in Viet Nam but mainly occupy the delta and the coastal areas while ethnic minorities mostly live in the highland areas, especially in central and northern Viet Nam (Figure 2). These maps highlight the spatial segregation between ethnic groups in Viet Nam’s 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 Viet Nam

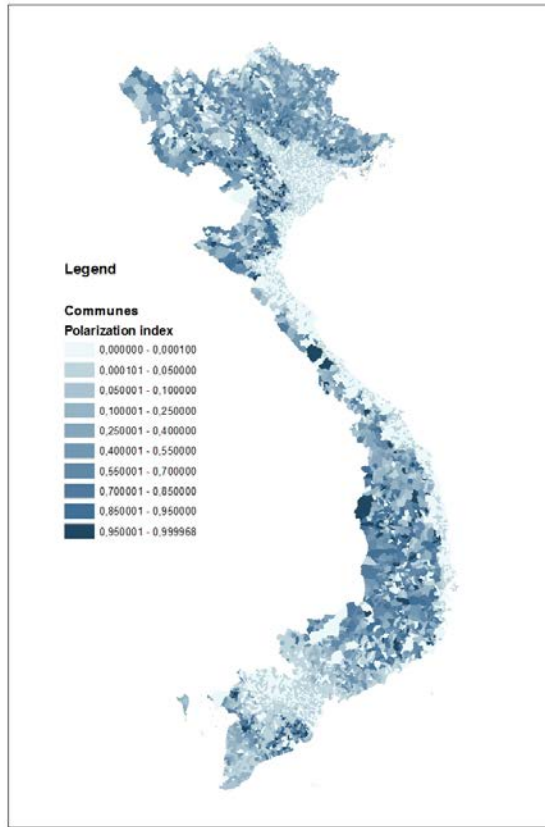
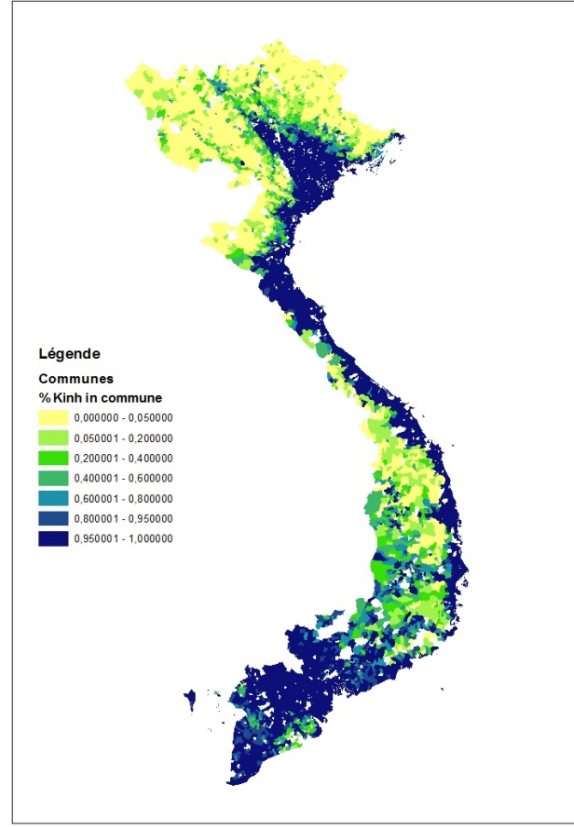


Figure 2. Kinh distribution in Viet Nam



Source : Authors' calculations based on 2009 Census and GeoQuery data.

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 (Montalvo 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 Montalvo and Reynal-Querol (2005) under the form of:

$$RQindex_c = 1 - \sum_{i=1}^N \left(\frac{0.5 - \pi_{ic}}{0.5} \right)^2 \cdot \pi_{ic}$$

Where π_{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 Viet Nam 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 Viet Nam, ethnic minorities often live in geographically difficult environments (i.e. mountainous). A recurrent question in studies on ethnic gap in Viet Nam 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, 2001). 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 Weerd, 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 Viet Nam, 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 (n=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 2008–16 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 (n=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 2008–16 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.¹⁰ Results are presented in another paper (Hoang and Saint-Macary¹¹, 2018). Two main conclusions can be driven from this vulnerability analysis: (i) vulnerability is high¹² 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 shocks¹³, 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

¹⁰ 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.

¹¹ Hoang, Q. and Saint-Macary, C. (2018). 'Vulnerability and ethnicity in rural Vietnam'. *mimeo* (working paper in progress but available upon request from the authors).

¹² 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).

¹³ 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.

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 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 2008–16 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.¹⁴ This reflects lower opportunities in mixed-communes to form diverse risk sharing network in terms of occupation than in non-mixed communes.

¹⁴ See section 3.3 for a presentation of this index.

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***
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 2008–16 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. δ_t are year dummies¹⁵, μ_d are district fixed effects, and

¹⁵ 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.

ϑ_{dt} , the interaction of both. ε_{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, 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.¹⁶ Indeed, there is evidence from the literature that social distance between ethnic groups may strongly vary

¹⁶ 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 interaction 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.

according to the relative size of each group.¹⁷ 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.

In all of the above models, the dichotomization of the ethnic (i.e. Kinh vs. non Kinh) may appear too simplistic, especially considering that Viet Nam 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 5.¹⁸ As expected, we observe that male and more educated household heads have larger networks.¹⁹ 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

¹⁷ 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.

¹⁸ Results from the Poisson regressions are available upon request from the authors.

¹⁹ 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.

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 Viet Nam. 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.

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 6.²⁰ 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

²⁰ 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.

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.²¹ 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 1C 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.

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

²¹ 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.

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 Viet Nam. 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 1C).²² 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.²³ 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 randomly in a commune have different occupations.²⁴ 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)).

²² 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.

²³ Because this measure is based on the name generators, the proportion is calculated among the three first contacts mentioned by the household head.

²⁴ More precisely, this index is built as a fragmentation index calculated as below:

$$FRAG_c = 1 - \sum_{i=1}^N \pi_{ic}^2$$

Where π_{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 Viet Nam standard industrial classification.

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 1C).

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 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 2). It could well be, 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

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Log of potential helpers	OLS		OLS		OLS		OLS		OLS		OLS		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)	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)	-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)	-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 2008–16 VARHS.

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

Ethnic similarity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Tobit	Tobit	Tobit	Tobit	Tobit	Tobit	Tobit	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 (0.040)	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)	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)	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)	-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 2008–16 VARHS.

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

Geographical similarity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	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 2008–16 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)	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)	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)	-0.001	(0.00)		
Night light from 1996 to 2000					-0.052**	(0.03)	-0.019	(0.03)	-0.022	(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 2008–16 VARHS.

6 Conclusion

In this paper, we have examined the differences in the risk sharing networks of Viet Nam's ethnic minorities and majority, as a way to approach the question of the persistence of the ethnic gap in Viet Nam. Exploiting a very rich rural household data set collected between 2008 and 2016 in several provinces of Viet Nam, 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 Hoang and Saint-Macary (2018) show that rural households in Viet Nam 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 Viet Nam, 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 Viet Nam'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 Viet Nam 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, A. and E. La Ferrara. 2005. 'Ethnic Diversity and Economic Performance.' *Journal of Economic Literature* 43(3): 762–800.
- Arcand, J.-L. and D. Jaimovich. 2014. 'Does ethnic diversity decrease economic interactions? Evidence from exchange networks in rural Gambia.' MPRA Paper 60497.
- Attanasio, O., A. Barr, J.C. Cardenas, G. Genicot and C. Meghir. 2012. 'Risk Pooling, Risk Preferences, and Social Networks.' *American Economic Journal: Applied Economics* 4(2): 134–67.
- Banerjee, A.V., and E. Duflo. 2005. 'Growth Theory through the Lens of Development Economics.' In P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*. Amsterdam: Elsevier.
- 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, C., and J. Charbonneau. 2012. 'How to Generate Personal Networks: Issues and Tolls for a Sociological Perspective.' HAL Working Paper 00512180. France: HAL Open Archive.
- Bramoullé, Y., and R. Kranton. 2007. 'Public goods in networks.' *Journal of Economic theory* 135(1): 478–94.
- Brandt, K., and F. Tarp. 2017. 'Characteristics of the VARHS Data and Other Data Sources', in Tarp, F. (ed). *Growth, Structural Transformation and Rural Change in Viet Nam: A Rising Dragon on the Move*. Oxford: Oxford University Press.
- Burt, R.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: Harvard University Press: 324.
- Carter, M.R., P.D. Little, T. Mogue, and W. Negatu. 2007. 'Poverty Traps and Natural Disasters in Ethiopia and Honduras.' *World Development* 35(5): 835–56.
- Chandrasekhar, A.G., C. Kinnan and H. 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, S., and C.B. Barrett. 2012. 'Social Network Capital, Economic Mobility and Poverty Traps.' *Journal of Economic Inequality* 10(3): 299–342.
- Coate, S., and M. Ravallion. 1993. 'Reciprocity without Commitment. Characterization and Performance of Informal Insurance Arrangements.' *Journal of Development Economics* 40(1): 1–24.
- Coleman, J.S. 1990. 'Social Capital in the Creation of Human Capital.' *American Journal of Sociology* 94 (Supplement): 59–120.
- Comola, M. 2010. 'The network structure of mutual support links: Evidence from rural Tanzania.' PSE Working Paper 2008-74. Paris: Paris School of Economics

- Currarini, S., M.O. Jackson and P. Pin. 2009. 'An Economic model of Friendship: Homophily, Minorities, and Segregation.' *Econometrica* 77(4):1003–45.
- De Weerd, J., and S. Dercon. 2006. 'Risk-Sharing Networks and Insurance against Illness.' *Journal of Development Economics* 81 (2): 337–56.
- Dercon, S. 2002. 'Income Risk, Coping Strategies and Safety Nets.' WIDER Working Paper 2002/22. Helsinki: UNU-WIDER.
- Déry, S. 2003. 'Distinctions ethniques et déforestation au Viêt 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: 231–60.
- Epprecht, M., D. Muller, 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): 349–68.
- Esteban, J., and D. Ray. 2011. 'Linking Conflict to Inequality and Polarization.' *American Economic Review* 101(4): 1345–74.
- Fafchamps, M. (ed.) 2003. *Rural Poverty, Risk and Development*. Cheltenham: Edward Elgar Publishing.
- . 2011. 'Risk Sharing between Households'. In J. Benhabib, A. Bisin, and M. Jackson (eds), *Handbook of Social Economics* 1 (1 B): 1255–79. Amsterdam: Elsevier.
- Fafchamps, M., and F. Gubert. 2007. 'The Formation of Risk Sharing Networks.' *Journal of Development Economics* 83 (2): 326–50.
- Fafchamps, M., C. Udry, and K. Czukas. 1998. 'Drought and Saving in West Africa: Are Livestock a Buffer Stock?' *Journal of Development Economics* 55 (2): 273–305.
- Fischer, I., T. Beuchelt, T. Dufhues, and G. Buchenrieder. 2010. 'Risk management networks of ethnic minorities in Viet Nam.' *Asia-Pacific Development Journal* 17(2): 93–118.
- Goldstein, M., A. De Janvry, and E. 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., A. BenYishay, and D. Runfolo. 2016. 'Overview of the geo Framework'. AidData. Available online at geo.aiddata.org.
- Grimard, F. 1997. 'Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire.' *Journal of Development Economics* 53(1997): 391–422.
- Habyarimana, J., M. Humphreys, D.N. Posner and J.M. Weinstein. 2007. 'Why Ethnic Diversity Undermine Public Goods Provision.' *American Political Science Review* 101(4): 709–25.
- Hardy, A. and Van Chinh Nguyen. 2004. 'Gérer la "question ethnique" sur les hautes terres : un exercice d'équilibriste'. In S. Doyon and B. De Treglodé (eds), *Viêt Nam contemporain*. Bangkok: IRASEC/Les Indes savantes.
- Hasegawa, M. 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., A. Storeygard, and D. Weil. 2012. 'Measuring economic growth from outer space.' *American Economic Review*, 102 (2), 994–1028.

- Hoang, Q. and C. Saint-Macary. 2018, forthcoming. 'Vulnerability and ethnicity in rural Vietnam'. Dial Working Paper, forthcoming. Paris: Université Paris-Dauphine.
- Hodler, R., and P. Raschky. 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., R. Gaiha, and W. Kang. 2011. 'Poverty, inequality and ethnic minorities in Vietnam.' *International Review of Applied Economics*, 25, 249–82
- Institute for Studies of Society, Economics and Environment (ISEE). 2011. 'Portrayal of Ethnic Minorities.' Printed Newspapers. Hanoi: The Gioi Publishers
- Jackson, M.O., B. Rogers, and Y. Zenou. 2016. 'The Economic Consequences of Social Network Structure.' IFN Working Paper No.116. Stockholm: Institute of Industrial Economics.
- Jaramillo, F., H. Kempt, and F. Moizeau. 2015. 'Heterogeneity and the formation of risk-sharing coalitions.' *Journal of Development Economics* 114: 79–96.
- Johny, J., B. Wichmann, and B.M. Swallow. 2017. 'Characterizing social networks and their effects on income diversification in rural Kerala, India.' *World Development* 94: 375–92.
- Karlan, D., R. Osei-Akoto, and C. Udry. 2014. 'Agricultural Decision after Relaxing Credit and Risk Constraints.' *The Quarterly Journal of Economics* 2 (129): 597–652.
- Khwaja, A. 2009. 'Can good projects succeed in bad communities.' *Journal of Public Economics* 93(7–8): 899–916.
- Ligon, E., and L. Schechter. 2003. 'Measuring vulnerability.' *Economic Journal*, 113(486): 95–102.
- McElwee, P. 2004. 'Becoming Socialist or Becoming Kinh? Government Policies for Ethnic Minorities in the Socialist Republic of Viet Nam'. In C.R. Duncan (ed.), *Civilizing the Margins : Southeast Asian Government Policies for the Development of Minorities*. Ithaca: Cornell University Press.
- McElwee, P. 2008. "'Blood Relatives" or Uneasy Neighbors? Kinh Migrant and Ethnic Minority Interactions in the Trùng 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*, 27: 415–44.
- Michaud, J. 2008. 'Flexibilité de l'économie chez les Hmong de la haute région du Viêt-nam septentrional'. *Aséanie* 27: 151–83.
- Miguel, E. and M.K. Gugerty. 2005. 'Ethnic diversity, social sanctions, and public goods in Kenya.' *Journal of Public Economics* 89: 2325–68.
- Montalvo, J.G., and M. Reynal-Querol. 2005. 'Ethnic Polarization, Potential Conflict, and Civil Wars.' *American Economic Review* 95(3): 796–816.
- Morduch, J. 1994. 'Poverty and vulnerability'. *American Economic Review*, 84(2): 221–25
- Morduch, J. 1995. 'Income Smoothing and Consumption Smoothing'. *Journal of Economic Perspectives*, 9(3): 103–14.
- Nguyen Hoa Thi Minh, 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: 145–62.
- Pasquier-Doumer, L., and F. Risso Brandon. 2015. 'Aspiration Failure: A Poverty Trap for Indigenous Children in Peru?'. *World Development* 72: 208–23.

- Rockenbach, 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.
- Saint-Macary, C. 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, S., and U. Beck. 2015. 'Ethnic Disadvantage In Viet Nam: Evidence Using Panel Data'. WIDER Working Paper 2015/097. Helsinki: UNU-WIDER.
- Townsend, R. 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, D., and D. Gunewardena. 2001. 'Sources 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 Assessment 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. List of additional tables

Table 1A. Descriptive statistics of households

Variable	Kinh	Non-Kinh	Difference	Kinh	Non-Kinh	Difference
	(n=8858)	(n=2317)	(T-test)	(n=5819)	(=2159)	(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 of 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 2008–16 VARHS.

Table 1B. 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.

Table 1C . Determinants of risk sharing network's size and similarity by ethnicity disaggregation

Dependent 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)	Tobit (9)	Tobit (10)	Tobit (11)	Tobit (12)	Tobit (13)	Tobit (14)	Tobit (15)	
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 2008–16 VARHS.

Appendix 2. 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 2A. 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 (0.003)	0.002 (0.003)	0.002 (0.003)	0.001 (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*** (0.019)	0.071*** (0.020)	0.071*** (0.020)	0.064*** (0.020)
HH size	0.002 (0.005)	0.005 (0.005)	0.005 (0.005)	0.003 (0.005)
Head completed primary school	0.027 (0.026)	0.032 (0.028)	0.031 (0.028)	0.032 (0.028)
Head completed lower secondary	0.045* (0.027)	0.042 (0.028)	0.042 (0.028)	0.045 (0.028)
Head completed upper secondary	0.118*** (0.031)	0.128*** (0.033)	0.129*** (0.033)	0.140*** (0.033)
Head is farmer	-0.014 (0.019)	-0.023 (0.019)	-0.023 (0.019)	-0.032 (0.020)
Head or Spouse is village native	-0.023 (0.030)	-0.031 (0.031)	-0.031 (0.031)	-0.030 (0.031)
Kinh	0.114** (0.048)	0.097* (0.052)	0.108** (0.051)	0.115** (0.051)
Asset value (log)	0.032*** (0.004)	0.034*** (0.004)	0.034*** (0.004)	0.036*** (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* (0.152)	-0.442** (0.175)	-0.337** (0.166)	-0.389** (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 2008–16 VARHS.