

Heterogeneity and the gender and ethnic earnings gaps in Vietnam

Quynh Hoang¹

IRD, UMR DIAL, 75010 Paris

PSL, Université Paris Dauphine, LEDa, 75016 Paris

François Roubaud

IRD, UMR DIAL, 75010 Paris

PSL, Université Paris Dauphine, LEDa, 75016 Paris

This draft: May 15 2016

Abstract

In this paper, we aim to assess the gender and ethnic earnings gaps in Vietnam by using the Labour Force Survey in 2007. We show that in general, ethnic earnings gap is higher than gender earnings gap and the gaps are very heterogeneous amongst different ethnic minority groups. Then, results from the multinomial logit model used to correct for selectivity indicate that the gaps depend significantly on the sector choice: the gender earnings gap is observed low in public sector while the ethnic earnings gap is widened in agriculture sector and reduces significantly in public and formal sectors. Quantile regression result suggests that the gender earnings gap is lowest at the bottom quantiles but does not vary much across the other quantiles while ethnic earning gap is higher at the lower quantiles of the distribution and getting narrower at the higher level of the outcome distribution. Our next aim is to assess how much of these gaps can be explained by individual characteristics and their allocation between different sectors. We use the well-known Oaxaca Blinder and Neumark method to decompose the gaps and the extended Appleton's full decomposition approach to take into account selectivity in sectoral allocation. Our findings show that differences in characteristics and sectoral allocation play an important role in explaining the ethnic earnings gap but it is not the case for gender earnings gap.

Key words: Vietnam, employment, gender earning gaps, ethnic earning gaps, earning equation

JEL classification: J31, J71, O53

¹ Corresponding author: quynhhoang176@gmail.com

1. Introduction

“Going inclusive” is the motto for growth path, emphasized and promoted by the OECD recently in order to cope up with the global issues of poverty, increasing unemployment rate, etc. Inclusive growth means that all people of the society are allowed to contribute to and benefit from economic growth. Over the last 20 years, Vietnam has obtained outstanding progress in fighting poverty; however, it is now facing this problem of “inclusiveness”. One of the key constraints for Vietnam to obtain inclusive growth is horizontal inequality or inequality among groups; in particular, between male and women; between ethnic minority and majority groups. Horizontal inequality, in many cases, is found to be persistent over a long period. When horizontal inequality persists, it could pull people in poverty trap and also lower other human development indexes. Then, the focus should be given on the disparity sources. Potential causes could be the differences in economic characteristics between males and females, ethnic minority and majority groups, however, given equal productive endowments and location, discrimination could be another explanation. In this study, we look at the gender and ethnic earning gaps in Vietnam and examine whether discrimination exists in the labor market.

Although less diverse than other countries in the region, Vietnam is characterized by a significant level of ethnic diversity. According to the last population Census (2009), 86% of Vietnam’s population is Kinh (the majority group), while around 14% belong to one of the 53 other ethnic groups officially registered in the country. Historically, ethnic minorities are mainly concentrated in mountainous and remote areas. While at the aggregate level, Vietnam records one of the best performances in the world in terms of economic growth and poverty reduction during the last two decades, ethnic minorities seem to have benefitted less than others to this ongoing process. Ethnic minority groups account for an increasing large share of the poor in Vietnam. In 2010, ethnic minorities represented only 14 percent of the population, but 47 percent of the poor and 66 percent of individuals in the poorest 10 percent of the population. The poverty rate was 66.3 percent among ethnic minority, compared to around 12.9% among the Kinh majority.

This diverging trend questioned the inclusiveness of growth pattern and the long term viability of such a model in terms of social cohesion. The issues at stake are why this pattern occurred, in spite of an increasing awareness of the problem by the Vietnamese government, and which type of policies can be implemented to reverse it. Numerous studies has been undertaken in the past years by independent academics or commissioned by the authorities. Our paper belongs to this strand of research, and focuses on assessing further gender and ethnic labor earnings gaps in Vietnam, taking advantage of a relatively comprehensive data in terms of ethnicity, the Labor Force Survey (LFS), conducted by the National Office of Statistics (GSO) in 2007 with the technical assistance by one of the authors.

Our paper is organized as follows: section 2 explores the literature review of the topic, section 3 describes the data, the reason why we choose this dataset in 2007 and based on descriptive analysis, we draw a broad picture and the main stylized facts concerning the gender and ethnic groups’ participation in the Vietnam’s labor market. Then, section 4 elaborates on econometric specifications to estimate the earning gaps, non-farm earning gaps, wage gaps and decomposition methods are discussed as well as strategies to tackle self-selection and endogeneity potential biases. The results from econometric estimations are going to be discussed in section 5 and section 6 concludes and draws some perspectives for further research.

2. Literature review

A long-standing literature on the economics of discrimination has indicated that there are two basic models of discrimination in the labor market: taste-based discrimination and statistical discrimination. The former is initiated by the work of Becker's (1957). In his model, discrimination is seen as a personal prejudice, or taste, against associating with a particular group or individuals prefer certain individuals or groups to others. Several possible sources of discrimination include employer, employee, or customer. There is no productivity-related attribute in the employing process; the workers from a less favored group have to accept lower wage or perform at higher productivity compared to the favored identical group. In the market, taste-based discrimination could be seen when a customer is willing to buy a good at a relatively higher price as a cost for his discrimination against some other sellers.

The second theoretical model of discrimination is proposed by Phelps (1972) and Arrow (1973) who focus on the statistical discrimination. Their assumption is that due to the asymmetric information, the employer cannot identify the ones who have better skills or some given qualifications. Thus, the firms or employers use easily observable characteristics such as race, ethnicity or gender to predict the expected productivity of the job applicants. In other words, based on stereotypes, the employers judge a prospective employee by their own imperfectly measured ability and performance as well as by that of their group (group average). Indeed, the literature on discrimination suggests that in the labor market, disadvantaged people in terms of race, gender and ethnicity are more likely to be treated unequally to the advantaged group and the earning gap can be seen as one of the consequences.

Empirically, there has been a rich literature showing that women have been treated unequally in the labor market and it is not a country-specific phenomenon. In a recent study, Nopo et al. (2011) use data of sixty four countries to provide a broad picture of the gender earnings gaps around the world, from developed to developing countries and across different regions. After controlling for observable socio-demographic and job characteristics, the gender earnings gaps vary between 8% to 48% of average women's earnings and it is worth emphasizing that the unexplained gaps are commonly observed for people having only part time job and low education level. The abundant studies on gender gaps not only examine the gaps but also determine the explanation factors for the gaps, such as differences in wage structure and women's labor supply (Blau and Kahn (2001), sectoral allocation (Tzannatos, 1999; Nordman et al., 2009; Nordman et al., 2011, etc.), liberalization policies in the labor market and institutional frame in each country (Weichselbaumer et al., 2007; Blau and Kahn, 2001; Cornish, 2007 and Tzanatos, 1999, etc.) while the unexplained parts of the earnings gaps could be due to unobservable factors and/or discrimination against women.

In Vietnam labor market, studies on the gender earnings gaps have been largely focused on the changes in women's participation in the labor force and inequality in wage between men and female workers. Gallup (2004) uses two rounds of Viet Nam Household Living Standards Survey (VHLSS) in 1992-93, 1997-98 and by estimating the wage equation, he finds that women in Vietnam earn much less than men with the same observable characteristics, although the differences in wages decrease between 1993 and 1998. Using the same source of data, Liu (2004a, 2004b) describes the wage structure development in Vietnam, emphasizing the gender wage gaps across sectors. She confirms Gallup's result that the Vietnamese women face lower wage. Furthermore, the gap is found highest in private sector, then State-owned enterprises (SOEs) and lowest in public sector. By decomposing the gap, she finds that although the gender pay gap is narrowed, discrimination against women in the labor market is higher and within sector differences contribute to the wage gaps more than between sector differences. In the same trend with Gallup (2004) and Liu (2004a, 2004b), Pham et al. (2007a) explore the VHLSS in 1993, 1998 and 2002 to investigate the gender wage gap by using mean and quantile

regression method. The result is that gender wage gap reduces by half over the period between 1993 and 2002 and most of the achievement is observed in the 1993 – 1998 period. The gap is found narrowing at most selected point of the conditional wage distribution and most sizeable at the top end of the conditional wage distribution.

Concerning the ethnic earnings gaps, studies on this topic are less abundant than gender earnings gaps due to the limited data sources and ethnic inequality is not the common topic for all countries. However, few studies worked on this topic in Vietnam are worth emphasizing as follows. Gallup's study (2004) reveals that non-Chinese ethnic minorities do not show lower wages in 1993, but they do show a 10 percent lower wage in 1998. Ethnic Chinese in Vietnam had a 25 percent wage premium in 1993, but this all disappeared by 1998. Using the VHLSS 2002 with Oaxaca decomposition method to decompose ethnic wage gap into treatment and endowment effects at both the mean and at selected quantiles of the conditional wage distribution, Pham et al. (2007b) find that there is an ethnic wage gap in the labour market but this gap is substantially narrower than the gap observed in another study using household living standard measures (Van de Walle and Gunewardena, 2001; Baulch et al., 2004). Their results from the quantile regression models reveal a gap of nearly 21 percentage points at the bottom decile compared to only four percentage points at the top deciles and the ethnic wage gap is mainly attributable to the differentials in returns to endowments. However, this study does not examine the ethnic wage gap for self-employed and the gap is considered at aggregated level between Kinh and Non-Kinh. More recently, Baulch and Vu (2011) carry out a mean decomposition using VHLSS of 2004 and 2010 to analyze the factors driving the ethnic expenditure gap. Their finding is that the ethnic expenditure gap increases from 58.1 percent in 2004 to 68.1 percent in 2010. Differences in endowments explain two fifth of the gaps in 2004 and nearly half of the gap in 2010 and it is noted that more than half of the expenditure gap is accounted for by differences in returns.

So far, the researches on the earning gaps in Vietnam, for both gender and ethnic gap, have focused on wage employment due to the lack and difficulties of collecting data on self-employment or agricultural sector, thus, the research results ask to be interpreted with caution because wage employment only accounts for a small fraction of the labor force in Vietnam. In the studies on gender pay gaps in the transition in Vietnam, Liu uses the traditional Oaxaca and Neumark decomposition as well as the Appleton et al. (1999) to correct for the differences attributed to sectoral choice differentials. However, she argues that her work focuses on the differences in returns of a wage-worker in one sector from another, thus, the two types of sample selection bias due to the fact that not everyone is working and the selection into wage employment, are dropped from the her analysis for simplicity. Thus, she uses a multinomial logit model (Lee, 1983) in the first step to correct for the selectivity between three different types of wage employment. Following Liu (2004a), Pham et al. (2007a,b) also use the above mentioned decomposition methods and use the Lee (1983) model to correct for selectivity when they decompose the gender and ethnic wage gaps at means and at selected quantiles of the wage distribution. And again, they do not provide analysis for the self-employed. In the case of Vietnam, ethnic minority people mostly locate in rural and remote area; farming and self-employed are their main work, thus, it would be not satisfied to estimate only the wage gap which might underestimate the real gap between the majority and minority. Similarly, regarding the gender gaps, if we only focus on the wage gaps, we could look at a small section of the labor market where women are likely to be overrepresented in public sector (Liu et al., 2004b), the gap might be underestimated. Then, our study will examine the gender and ethnic earnings gaps across different employment sectors. Moreover, our study sheds a light on this research topic by not only considering all ethnic minority groups as a whole against majority but going further by disaggregating ethnic minority groups. It could be the case that people from one ethnic minority group could earn equally or even higher income than the majority, which we could not realize when the whole 53 ethnic minority

groups are aggregated. In this paper, we will start analyzing the gender and ethnic earnings gaps, non-farm earnings gaps, wage gaps, then, decomposing the gaps in different employment sectors as well as correcting for selectivity bias in our study. It is worth emphasizing that the ethnic group classification will be considered under both ways: ethnic minority group aggregation and disaggregation.

3. Data and descriptive statistics

3.1. The data and ethnic group's aggregation

Our study relies on the Vietnam Labor Force Survey 2007 (LFS2007) which is characterized by a particular sampling scheme that makes it valuable for our study focus. Compared with previous studies, the large LFS's sample size (more than 173,000 households) provides more precise estimates at a higher level of inference (provincial vs. regional). It allows us to get a higher level of ethnic disaggregation, to reflect better ethnic heterogeneity in Vietnam. Indeed, this is also an advantage of the LFS2007 in comparison with the following LFSs which do not contain sufficient information on ethnic group disaggregation (Tay, Thai, Muong, Kho-me, Nung, Hmong and the rest). Most of the previous quantitative works on ethnic issues in Vietnam have focused on the major separation between Kinh and Non-Kinh (Van de Walle, 2001; Baulch and al., 2004). More recent papers have gone further in considering 8 ethnic groups: Kinh & Hoa, Khmer & Cham, Thai, Tay, Muong and Nung, Other Northern minorities, Central Highland minorities, Others (Baulch et al., 2007, 2008, and 2010). Taking advantage of a much higher sample size (173,000 households, compared to less than 10,000 households in previous studies), we are able to relax the data constraint, by increasing the number of categories with statistical significance. This point is worth highlighting as within differences between ethnic groups may be quite substantial, and in some cases, probably higher than the ones observed between aggregate categories. The second main strongpoint of our work is that for the first time ethnic identification has been asked to all members of the households, while previous surveys only collected data for the head of household, relying on the implicit hypothesis that all households are endogamous. We are also able to investigate the earnings gaps in the informal sector and self-employed when the previous analysis had to be confined to only wage workers due to the limited data.

Vietnam's official classification of ethnic minorities considers 54 different ethnic groups. According to the LFS2007 (Table 1), 86% of the population is Viet or Kinh (the majority group) and 14% belong to ethnic minorities, representing 11.2 million individuals. Among the 53 other ethnic groups, the distribution is highly skewed (evenly distributed). Some groups account for more than one million (Thai, Tay, Muong, Khmer) or near one million (Nung, Hoa), while others represent a few thousands (Pu Peo, Bo Y, Co Lao, etc.) or even a few hundred people (Brau, Si La).

3.2. Gender and Ethnic minorities in the labour market: a descriptive analysis

The main labor market indicators computed from the LFS2007 offer evidences of strong differences Kinh and ethnic minorities regarding the ethnic labor market attachment but some difference between males and females (Table 2). First, the working age population (over 14 years old) share is higher for women than men (76.1% vs. 72.9%) but the employment rate is observed higher for men by 8%. Women are more likely to have a second job than men, which reflect the fact that women might choose to have a temporary job, attached to farming activities in order to have time taking care of their children and family. In terms of ethnicity, the participation rate is significantly higher for Kinh (76% vs. 69%), reflecting higher fertility and mortality rates (Friedman et al., 2001), proportion of children, and dependency ratio among ethnic minorities. Hoa are the only ethnic group to have a

higher proportion of working age population than Kinh.² Conversely, the activity rate is in average almost 10 percentage points higher for ethnic minorities (83% vs. 73%), due to lower school attendance (for the young generation) and a higher rate of households' labor force mobilization in primary sectors. Consistently, the rate of unemployment is lower for ethnic minorities (1% vs. 2%), comforting the hypothesis of luxury unemployment on Vietnamese labor market stressed in other studies (Razafindrakoto et al., 2011), while the rate of multi-activity (holding a second job) higher (23% vs. 17%), reflecting a higher attachment of ethnic minorities to farming activities. Surprisingly, time related underemployment rate³ in the main job does not differ significantly between majority and minorities (around 11% for both), in spite of lower working hours for the latter. This feature probably results from a lower propensity to declare wanting more jobs within ethnic minorities, because of scarce labor market opportunities.

For this set of indicators, Kinh's situation contrasts sharply with all other ethnic minority groups (except Hoa), even if differences can be observed within them. Hmong are by far in the worst position, with only 55% in the working age bracket, compensated by a rate of activity as high as 92%. Unemployment is quasi inexistent while 35% of the working population are constrained to hold a second job in order to earn a living. Hoa community, occupying the other end of the minorities' ladder, presents exactly the reverse figures, with the highest rates of working age population and unemployment (80% and 4% respectively), the lowest rates of activity (65%) and multi-activity (6%), and one of the lowest underemployment rate (6%).

If we now turn to the characteristics of the labor force by gender and ethnic groups (Table 3). It is seen that women and men share somewhat similar proportion of locating in rural and urban areas, similar age levels but women in the labor force has lower years of schooling than men (7.8 vs. 7.4) and proportion of married women is also lower. Ethnic minorities are systemically more often located in rural areas than Kinh (90% vs. 73%; Table 3). The only noticeable exception is Hoa, where 78% are urban. In fact, for most of the ethnic groups, the proportion of rural labor force exceeds 95%, except for Gia-rai (94%), Nùng and Other Central Highland minorities (93%), Tay (92%), Khmer and Cham (89%), and E De (87 %). If the women share doesn't vary a lot around 50% (not taking into account Hoa, where 58 % are male), ethnic minorities are always younger (35 years old in average vs. 39 for the Kinh), and less educated. Kinh labour force has in average 8.5 years of achieved schooling, while ethnic minorities have only 5.8 years. A fourth of Kinh labour force has completed higher secondary education, but they are only one out of ten in ethnic minorities. 8 ethnic groups are below 5%. Again, Hmong are in the worst situation (with 2 years of schooling and 1% attended tertiary education). Tay, Muong and to a lesser extent, Hoa perform relatively well in this respect (8.1, 8.0 and 7.5 years of schooling each).

The structure of jobs by institutional sectors gives further insights to job allocation between men and women and among ethnic groups. Men present more than women in public and formal sector while there is no difference between these groups when they are engaged in the informal sector. Agriculture is the only sector where women are involved more than men. Obviously, the massive participation of ethnic minorities in agricultural activities leads to a systematic underrepresentation in other institutional sectors. Nevertheless, the remaining structure is unevenly

² The specificity of Hoa's group, which we will encounter for almost all labor market indicators, leads some authors to prefer considering Kinh and Hoa together, and differentiate them from other ethnic groups. In this paper, we choose to identify Kinh and Hoa separately. As Hoa are on most fronts well off than the rest of the population (including Kinh), one has to remind that differences between Kinh and non Kinh (excluding Hoa) presented in the tables should be systematically higher, in favour of the Kinh.

³ Time related underemployment the ration of the occupied population working less than 35 hours during the week of reference and wanting more hours.

distributed. In broad terms, ethnic minorities are relatively more underrepresented in the nonfarm private sector than in the public sector (Table 2). This feature is due to a more active inclusion policy in the latter. Notwithstanding, the share of public jobs for ethnic minorities is more than twice lower than Kinh's one (5% vs. 12%). Within the private sector, ethnic minorities are engaged in formal household business (3% vs. 9% for Kinh), followed by the informal sector⁴ (7% vs. 26%) and domestic enterprises (2% vs. 6%). In this respect, the foreign enterprises perform far the worst: the relative chance for Kinh to get a job in foreign enterprises is more than five times higher than for ethnic minorities (2.2% vs. 0.4%), and twelve of our nineteen ethnic groups have no access to this sector. Such a result is highly problematic, as foreign enterprises have been both the fastest growing sector during the last decade and the one which provides the best jobs in terms of labor conditions. The exclusion of ethnic minorities from foreign enterprises is probably more due to adverse location, given the high concentration of FDI (85% in Hanoi and Ho Chi Minh City surroundings; ADB, 2004; Cling et al., 2009), and limited skills availability among ethnic minorities (Foreign enterprises recruit mainly semi-skilled workers) than to explicit discrimination.

Going into more details, evidences suggest that some ethnic groups seem to be keener to work in specific institutional sectors. This is for example the case of Hoa's labor force, with a strong predilection for formal household businesses where 30% are engaged, but only 8% of Kinh and 3% of non Kinh. Khmer and Cham seem to privilege the informal sector, while Central Highland minorities are globally excluded from the formal private sector. Tay are the best integrated in the public sector and administration (9%), apart from Kinh (12%), while the probability to get a public job for Co-ho, Ra-glai and Hmong is respectively 7, 8 and 12 times lower than Kinh's one.

Due to their sectoral allocation, the jobs' profile of ethnic minorities differs substantially from Kinh's ones. The predominance of farming activities in small family businesses has notable consequences on the key characteristics attached to jobs. The percentage of wage workers is more than twice lower (14% vs. 33%), substituted by own account and unpaid family worker's status. Accordingly, all kinds of benefits provided by the job's attachment are much less defunded, leading to inferior job's quality. Only 7% of the ethnic minority labor force are covered by written labor contracts (19% for Kinh), 6% benefit from official social protection schemes, 6% get public holidays and 7% pay slips (vs. 16%, 17% and 19% respectively). Even restricted to wage workers, ethnic minorities are significantly less protected than their Kinh's counterparts, in part because they work more often in institutional sectors where labor conditions are more precarious. With 15 years, tenure is higher for ethnic minorities, reflecting a lower job's mobility (12 years for Kinh), while they are working less hours. Here again, the job's situation is worst for Central Highland minorities than for Northern ones, while Hoa are in some respects relatively privileged even compared to Kinh. In accordance with their massive participation in subsistence farming activities, Hmong are the ones with the most unfavorable integration in the labor market. However, to go beyond these unconditional averages, one should take into account which part of these features can be explained by human capital differentials.

⁴ According to GSO, the informal sector comprises all non-registered (business license), non-farm household businesses. Formal household businesses are the registered ones.

4. Methodology

4.1. Earnings determinations

Mincer earning equations

In this paper, we estimate the classical Mincerian earnings equations, with different specification to address potential self-selection and endogeneity biases, and check for robustness of the results. Our simplest specification is as follows:

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

where $\ln w_i$ is the natural logarithm of the observed hourly earnings for individual i , x_i is a vector of observed characteristics, β is a vector of coefficients and ε_i is a disturbance term with an expected value of zero. The dependent variable is the log of real hourly earnings. In the pooled regression, we use a dummy for gender and dummies for ethnicity. In our regression, independent variables include productivity-related factors such as education and market experiences. Due to the data's nature, we do not have information about individual's actual experience but merely potential experience which is measured by the time that the individuals have spent in the labor force. In order to account for differences in consumer price levels by location, we use the deflators drawn from the VHLSS2006, which distinguish 16 domains: the 8 official regions by area (urban/rural). The first regression is a basic Mincerian equation, including sex, education and potential experience. Second regression includes province fixed effects (64 provinces), to partially tackle endogeneity concerns and capture non-price local effects. In the third regression, we keep the province fixed effect but add independent variables of the institutional sector and the urban/rural divide. In terms of ethnic divide, models are estimated at the usual dichotomous level (majority vs. minorities) and using the full disaggregation of ethnic minorities.

Multinomial logit model

Traditionally, studies on earning gaps use a simple OLS regression for the earning estimation where only actively occupied people are considered. However, if a significant part of the sample is not working, endogenous selection might be an issue if unobservable factors in the earning equation and in the participation equation are correlated. Besides the selection bias as mentioned above, another source of bias arises in running OLS regression is the selective decision of people involved in different sectors. Thus, using a binomial selection procedure might not be sufficient to correct for all relevant distinguished market outcomes in a labor market of a developing country like Vietnam. Indeed, besides the fact that earnings are only observed for people who work, people who choose to engage in public sector have different characteristics than people who engage in formal, informal and agricultural sector and those characteristics could affect their earning gaps. Therefore, we address these two potential biases by using the multinomial selection model presented in Bourguignon et al. (2007, hereafter BFG), relying on a generalized version of the Dubin-McFadden correction⁵.

The multinomial selection model is given as:

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

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

Where $s=1, 2, 3, 4, 5$, reflecting 5 alternatives that each individual can choose: i) public sector

⁵ See Bourguignon, Fournier and Gurgand (2007) for a survey of the available methods to obtain consistent estimates of β_s and γ_s with two-step procedure

employment; ii) formal sector employment; iii) informal sector employment; iv) agricultural sector and v) not working people. The vector Z represents the maximum set of explanatory variables for all alternatives and the vector X contains all determinants of the earnings; y_s is the earning corresponding to each employment choice presented by y_s^* - a discrete-choice variable. In the first stage, a multinomial logit model estimating the predicted probabilities of each individual's participation in sector s is used to calculate the correction term. Then, we estimate the earning equation including the correction terms for each alternative of sector choice. Ideally, we need an instrument which affects only the propensity to work but not correlated with the wage/earnings. Due to the data constraint, the first stage selection equation uses a set of individual characteristics which are assumed not to influence the earning levels, including: relationship to the household head, household size and dependent ratio (number of non-working age individuals divided by the total number of individuals in the household). Besides, other independent variables used in the second stage (earning equation) are also used in the first stage, which are gender, marital status, ethnic identity, years of schooling, potential experience and rural area.

Quantile regression

The OLS regression above represents earning gaps at the mean, however, the findings hold under the assumption that possible differences in terms of the impact of the exogenous variables along the conditional distribution are unimportant. In fact, the gender and ethnic gaps might vary across the outcome distribution, especially at the lower and higher tails of the earning distribution, therefore, we rely on a quantile earnings regressions which consider specific parts of the conditional distribution of the hourly earnings and indicate the influence of the different explanatory variables on conditional earnings respectively at the bottom, at the median and at the top of the distribution. The model is as follows:

$$Q_\theta(\ln w_{ij} | x_{ij}) = \beta_{ij,\theta} x_{ij,\theta} + \alpha_{j,\theta} k_{j,\theta} \quad (4)$$

Where $Q_\theta(\ln w_{ij} | x_{ij})$ is the conditional quantile function, $x_{ij,\theta}$ represents the set of characteristics for each individual i in ethnic group j and $k_{j,\theta}$ is the dummy for each ethnic group j in each quantile θ^{th} . Thus, the set of $\beta_{ij,\theta}$ and $\alpha_{j,\theta}$ coefficients provide the estimated rates of return to the different covariates at the θ^{th} of the log earning distribution.

4.2. Earnings gaps decomposition

To start with the decomposition part, our study uses the Oaxaca and Blinder decomposition method. We estimate separately the earning equation (1) for Kinh and non-Kinh/ethnic minorities (aggregated together) and then use the Oaxaca-Blinder decomposition techniques to identify the sources of ethnic earning gaps. Similar process is carried out to decompose the gender earning gaps where Kinh is replaced by male and non-Kinh is replaced by female group. In this approach, the decomposition is then given by:

$$\overline{\ln w_k} - \overline{\ln w_n} = \beta_k (\bar{x}_k - \bar{x}_n) + (\beta_k - \beta_n) \bar{x}_n \quad (5)$$

where w_k and w_n are the average earnings for Kinh (male) and ethnic minorities (female), respectively; \bar{x}_k and \bar{x}_n vectors of average observed characteristics for Kinh (male) and minorities (female); β_k and β_n the estimated coefficients. The first term on the right side of the equation measures the earnings gap due to characteristics (the "explained" part). The second one corresponds to differences in returns between the two groups, which may come from unobservable characteristics and/or discrimination ("unexplained" part or residual). This method, however, is under an argument of index problem, in which, we need to determine which earning structure is nondiscriminatory. If we choose the high-earning group's (ethnic majority or males) coefficient as non-discriminatory, the low-

earning groups (ethnic minority or females) are assumed to be underpaid or faced discriminated. Meanwhile, if the low-earning group's coefficients are considered as non-discriminatory benchmark, the other groups are supposed to receive nepotism. The literature has offered different approaches to address this issue. The method used in our study is Neumark's approach where β^* is introduced as a non-discriminatory coefficient by estimating the pooled sample using the weighted average of the earning structure of the high earning and low earning groups.

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

The first term is the earning gap attributable to differences in characteristics while the second and the third terms represent the earning differential between the actual and pooled returns to characteristics for the two groups. However, this method does not take the sector choice between high earning group and low earning group into account, which is argued as being a determinant affecting the earning gap. To overcome this issue, Appleton et al. (1999) proposes a full decomposition method which takes into account differences in sectoral choice between the two groups and still follow the Neumark's approach to avoid the index problem. The full decomposition starts with the idea of considering ethnic majority (male) and minority (female)'s earning as the sum of sectoral earning weighted by the proportion of workers in each sector. Let \bar{p}_{ks} and \bar{p}_{ns} be the sample proportion of Kinh (male) and non-Kinh (female) participating in each sector s . Thus, their earnings can be expressed as:

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

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

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

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

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

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

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

5. Results

5.1. Earnings determinants

5.1.1. Mincer equation estimation

The pooled OLS regression following equation (1) is presented in table 4. As expected, the goodness of fit is similar to what is usually obtained in the literature. It increases from model 1 to model 3 (adjusted R^2 improving from 0.14 to 0.26). Gender and human capital operate according to the literature and are highly significant: women suffer a penalty between -15% and -17% compared to their male counterparts, depending on the specifications (Nguyen et al., 2011). Returns to education is convex as in many other studies in developing countries, and returns to potential experience is concave (but increasing with the professional cycle as the maximum is achieved after more than 30 years of experience). The institutional sectors are ordered according to the expectations: from agriculture – the lowest rewarding sector, followed by informal sector, formal private sector and at the top of the pay ladder, public sector and foreign enterprises. Finally, location matters: province fixed effect are at play, suggesting that local labour markets are not fully integrated, while jobs located in rural areas get in average and *ceteris paribus*, a slight earning premium (4%), once lower prices level are accounted for.

Regarding the ethnicity, in the first sets of models, which consider only the binary partition between Kinh and Non-Kinh, the coefficient is highly significant, the latter perceiving between -35% (model 1) and -22% (model 3) than the former, conditionally to the independent variables included. At this stage, two provisional conclusions can be drawn from these results. First, it seems that in Vietnam, ethnic gap is more a matter of concern than gender gap in the labor market. Second, there is a strong presumption that ethnic minorities are suffering from "poor" location (concentration in remote areas) and even more important job segregation. Just taking into account institutional sectors reduces the ethnic gap by 10 percentage points (from -32% to -22%).

If one disaggregates Non-Kinh into different ethnic groups, to better reflect ethnic diversity, the general pattern is that Kinh earns more than their fellow citizens. 42 of the 57 estimated coefficients (19 ethnic groups, 3 specifications) are significantly negative. In the simple mincerian equation (model 1), only Hoa, Khmer and Ede get (conditional) have higher retributions, while Ba Na and Cham gain as much as Kinh. In the full model 3, Ede, Ba Na, Khmer and the residual Other minorities are in line with Kinh, Co-ho and Cham being the only to get more. The coefficients range is extremely heterogeneous, suggesting huge differences of situation between ethnic groups: from 0.28*** (Ede) to -1.42*** (Hmong) in model 1; from 0.14*** (Co-ho) to -1.15*** (Hmong) in model 2; and from 0.28*** (Co-ho) to -0.94*** (Hmong) in model 3. Consistently with previous findings, Hmong is the most disadvantaged group. All in all, these results confirm the need to differentiate between ethnic minorities for analytical and policy purpose.

To investigate further the gender and ethnic earning gap, we restrict our sample to nonfarm labor income (Table 5). Different reasons can be suggested to justify such an option. First, agricultural incomes are among the most difficult to capture through a single question, like in classical labor force surveys. Second, agriculture labor market is known to be functioning very differently from nonfarm sectors one, which implies that earnings functions should be sector specific. Third, as we saw that job segregation is probably at play, and ethnic minorities are concentrated in farming activities, it is all the more interesting to focus on what is happening outside agriculture. Finally, with more than 10,000 ethnic minorities observations (whether for nonfarm workers or for wage workers, we able to go beyond previous studies which had to focus only on rural issues).

Obviously, running regressions on nonfarm paid workers raise an issue of selection bias: ethnic

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

Findings from table 5 show that nonfarm ethnic gap is highly reduced. At the aggregate level, the gap, though still significant, ranges from -6% (model 1) to -3% (models 2 and 3). Compared to the gender gap, the ethnic gap is negligible: gender gap remains at nearly 16% against women for nonfarm jobs, and 10.5% for female wage workers. Once the ethnic groups disaggregated, many of them earn as much or even more than Kinh. Only 13 of the 57 estimated coefficients are significant at 1% level. The highest penalty is suffered by Ra-glai (-0.42*** in model 1 and -0.36*** in model 2) for nonfarm jobs, and a few Central Highland minorities (Ê Đê, Ba na and Other).

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

5.1.2. Mincer equation estimation with selectivity correction

In this section, our results are reported in two parts. The first part (Table 6a, b) describes the marginal effects from the multinomial logit model and second part (Table 7) highlights the results from earning equation estimation, in which the selection bias is corrected by a multinomial logit model.

The marginal effects from the multinomial logit estimation/first stage equation show us the determinants of the allocation of people into different sectors. As expected, we observe positive effect of education in public sector, formal and informal sector. The effect is strongest in public sector, convex in formal sector and concave in informal sector while it has negative effect on allocation into agriculture. As expected, experience also has positive marginal effect on selection into different sectors and negative effect on being inactive. People coming from larger household have lower probability working in public sector, informal sector and agriculture sector while they have higher probability to work in formal sector and being inactive. People are less likely to work in public and formal sectors when having high dependence ratio. The opposite effect is observed in informal and agriculture sector.

Regarding our variable of interest, women are more likely to work in public and formal sectors than men, keeping anything else equal. Ethnic minority has higher probability of working in public sector and agriculture sector while they have lower probability of working in formal and informal sector. In public sector, compared to Kinh, Tay, Nung, Hmong, Gia Rai, Ra-glai, Co-ho, other Northern minorities and other central highland groups and other groups have higher probability of working there while Hoa is less likely to work in public sector than Kinh. Regarding formal and informal sectors, there is almost no significant difference of allocation among different ethnic groups and as expected, almost all ethnic minorities groups have higher probability of engaging in agriculture than Kinh,

except for Hoa.

Earning equation estimates

After correcting for selectivity into different sectors, Table 7 presents the earning equation estimations. It is seen that returns to education are positive and significant in all cases but the link is convex in public sector and concave in informal and agriculture sector. We also find positive impact of potential experience on earnings in all sectors but the patterns are slightly different, which are concave in public and informal sector but convex in formal sector.

Although women have higher probability of some sectors, they face a lower earning gap in all sectors except in agriculture after the sector selection is controlled for. All else equal, women earns 7.7%, 10.6% and 18.5% less in public, formal and informal sectors, compared to men. Meanwhile, the ethnic earning gaps vary considerably among the sector. Being an ethnic minority, an individual can get an earning premium of 2% higher than their Kinh counterparts if they work in public sector, however, their earning will be 18% less when they are engaged in formal and informal employment. Moreover, we find a significant high gap in earning between Kinh and non-Kinh in agriculture sector, which is 48 percentage points.

When the earning equations are estimated with disaggregated ethnic groups, we observe again that the coefficient is extremely heterogeneous between ethnic minority groups. In public sector, Thai, Tay and Giay earn more than Kinh while Muong, Khmer, Ede and Ra-glai earn less. In the formal sector, except for Hoa, the coefficient is negative and significant for almost all ethnic minority groups, mostly around -0.2 but in the extreme case, we find a coefficient of -1.1 for Hmong group which is only significant at 10% level. In the informal sector, we also find very large range in the coefficients, from -0.1 (Muong, Khmer, Ba Na) to -0.7 (Hmong) and -0.8 (Other Central Highland minorities). In line with our discussion above, the ethnic earning gap is huge when people work in agriculture sector and the gap varies from -0.2 (Cham); -0.4 (Khmer); -0.5 (Tay), etc. to -0.9 (Dao); -1.1 (Gia Rai, Ra-glai) and -1.3 (Hmong). That means everything else equal, Gia Rai, Ra-glai and Hmong earn less than Kinh counterpart by 66 to 69 percentage point. Through all different sectors, it is notable that Hmong, Gia Rai and Ra-glai are the most disadvantaged ethnic groups.

In short, to some extent, these results confirm our strong presumption that ethnic gap is more a matter of concern than gender gap in Vietnam labour market, especially the heterogeneous earning gaps among the ethnic minority groups.

5.1.3. Quantile regression

In this part, we estimate the regressions at five quantiles of the log earning distribution, namely 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles. Results from table 8 show that gender earnings gaps do not vary much across different quantiles while the ethnic earnings gaps are higher at the lower quantiles of the distribution and getting narrower at the higher level of the outcome distribution. Besides, we can see clearly the heterogeneity with the ethnic minorities when we estimate at different section of the earning distribution. For instance, participating in the low-paying job (at the 10th quantile, Hoa people suffer a penalty of 12.4% but if they are involved in the highly-paying job (90th quantile), they would gain 7.2% higher than Kinh. For the ethnic groups which suffer most in the earning gaps, the coefficients are higher at the first two quantiles of the distributions, such as Hmong, Gia rai, Ra-glai, Nung, Central Highland, etc, this result is in line with other studies in developing countries where the earning gap is high at low quantiles of the earning distribution. Again, we also observe the heterogeneity in earning gap across different quantiles amongst ethnic minority groups.

5.2. Gender and ethnic earnings gaps decomposition

Neumark decomposition of gender and ethnic earnings gaps

In this section, gender and ethnic earnings gaps that we find above are decomposed by using Neumark approach. Regarding the ethnic earning gaps, we not only decompose the gap between Kinh and non-Kinh but also extend our analysis by decomposing the gap amongst different ethnic groups. This is motivated by the OLS result above that the earning gaps are very heterogeneous once we disaggregate the ethnic minority groups. Moreover, there might be a case that some given ethnic groups are more discriminated than other groups, thus, we will decompose the earning gaps by pair, each ethnic group along the row (horizontal axis) is compared to other group along the column (vertical axis), thus, positive coefficient means that group on the row has higher earnings than the group on the column. Due to the high number of ethnic groups but limited observations in small size groups, we only choose 10 largest ethnic minority to examine the earnings gaps between them and Kinh and amongst themselves. The result is gathered in a matrix in terms of the raw gaps and the unexplained proportion.

Similar to what we have seen in the last section, the raw gender earning gap is relatively large but not as high as the ethnic earning gaps, which are 0.224 and 0.526 respectively (Table 9). When the gaps are decomposed, we find that the differences in individual characteristics such as years of education, work experience and geographic variables could explain only 15.6% of the gender earning gaps while 84.4% of the gaps are left unexplained. For the ethnic earnings gaps, the differences in endowments account for nearly 61% of the gaps, which is expected due to the huge difference in education level and living location between the Kinh and non-Kinh groups.

Once we disaggregate the ethnic minority groups, the gaps and the proportion of unexplained part vary remarkably as presented in table 10. When the earning gaps are decomposed between Kinh and each of the ethnic minority groups, we find that Hoa and Ede perform better than Kinh and the endowment can explain for about 62% and 40% of the gap, respectively. Amongst 5 largest ethnic minority groups (Tay, Thai, Muong, Khmer, Mong), the gaps between them and Kinh are quite similar, around 0.5 with half of it unexplained, except for Khmer and Mong. The earning gap between Khmer and Kinh is much narrow, at 0.079 while the gap is significantly large between Kinh and Hmong, which is 1.5. It is also noted that Hmong lags behind all other groups and one of the reason is education as Hmong is among the least developed group in terms of education and training (Phung et al., 2014). Meanwhile, Ede, another ethnic minority group has earned higher earnings than all other groups. This matrix also helps us to see that there is no common pattern of ethnic earnings gaps between each ethnic minority group and the ethnic majority group as well as the unexplained part in the gaps which fluctuate from almost zero to 117%.

When we look at the decomposition by sector (Table 11), we find out that the huge earning gaps are attributed greatly to the differences between sectors. In public sector, both gender and ethnic earnings gaps are very small, at 0.02 and -0.07, respectively. It is also noted that the ethnic earnings gaps becomes negative, which means that ethnic minority receives higher earnings than Kinh and 14.6% of this gap are left unexplained. Regarding the gender earning gaps, they are relatively similar in private, informal and agriculture sectors while regarding the ethnic earning gaps, it is very low in private and informal sectors but extremely high in agriculture. The finding of gender gap in private sector is consistent with other previous studies on Vietnam labor market (Liu 2004a; Pham et al., 2007a). As the State owned enterprises were downsized, women who overrepresented in this sector became redundant and faced the risk of higher pay gap because working in the SOEs is more likely to be protected than in the private sector (Liu, 2004b). Although other studies on gender earning gaps in Vietnam have not explored the gaps in informal sector, our finding of high gender earning gaps in this

sector is found in line with other studies in African countries (Nordman et al., 2011). Since the differences in observable characteristics could not explain much for this gap, other factors might be involved. It is likely that women work in a lower paid segment of the informal sector since the higher paid jobs are more physical demanding. Indeed, manufacturing & construction accounts for the highest proportion of jobs (43%) in the informal sector (Cling et al., 2011), thus, women might be disadvantaged in this employment segment. Besides, it is possible that women choose to work in lower productivity jobs in order to have more free time to take care of their children and housework. This also explains the relative high gender earning gap in agriculture too as in Vietnam, men in rural areas are better paid than most women because they take on different (heavier and more dangerous) tasks (WB 2012). Another possible explanation could be considered is that in the informal sector, people need physical capital to start their own small business such as street vendor, then, the earning gap is observed because women has lower capital than men.

As the previous studies on gender gaps lack the informal sector and self-employed, it is interesting to look further at the subgroup of wage earners and non wage-earners (including self-employed and family workers. The decomposition is presented in table 11b. It is seen that in the private sector, the gender gap is slightly narrower for wage workers (0.24) compared to non-wage workers while in the informal sector, the gap for wage worker doubles the gap for non-wage workers. In all cases, the unexplained part accounts for more than 80% of the gaps. These results reinforce the idea that in the informal sector, the employers can set the wage by themselves, leaving the room for discrimination.

Full decomposition of gender and ethnic earnings gaps

Being a developing country, Vietnam's labor market comprises of four main types of employment sectors which are public, formal, informal and agriculture sectors. Those four sectors have their own characteristics with basic differences in job seasonality, wage structures, etc. which potentially affect the gaps in earning between different groups. As presented in section 4, we will decompose the earnings gaps taking into account the allocation of each group (male and female or ethnic majority and minority) in each of the four sectors. The full decomposition of gender earnings gaps are reported in table 12.

It is seen that within-sector differences in earnings explain for 96.3% of the gender gap, of which, 26.1% is attributed to differences in characteristics between male and female. The positive signs at high magnitude on the deviations in male and female returns imply that men receive nepotism while women face discrimination in their job within sector. On the other hand, although the differences in sectoral allocation between these two groups account for only 3.6% of the gender earnings gaps, most of the differences are explained by differentials in characteristics. The negative signs of the difference between the actual and predicted sectors for men and women help to narrow the gender earning gaps attributed to difference between sectoral allocation. This is probably due to the fact that women account for a high proportion in the public sector where the wage is higher and more stable (Liu 2004a). After correcting for the selectivity into different sectors, the gender earnings gap decreases noticeably but pattern of the decomposition does not change much.

A full decomposition of the ethnic earnings gap is presented in table 13. On the contrary to gender earnings gap, about two third of the ethnic earnings gap is due to the differences in sectoral allocation between ethnic majority and minority groups. Of which, the differences in characteristics that generate the different sector choice mostly explain the gaps. This finding corresponds to what we have seen earlier that the ethnic earnings gaps are captured almost entirely in agriculture sector. More than half of the ethnic earnings gaps could be reduced if ethnic majority and ethnic minority groups are allocated equally in the labor market. Within sector, the differences in characteristics explain half of

the gap; the other main contribution to the within sector gap is the difference in actual and neutral returns to ethnic minority characteristics. On the contrary to gender earning gaps, we find even larger ethnic earnings gaps after correcting for sector selectivity.

6. Conclusion

This paper uses the Labor Force Survey data in 2007 to estimate and analyze the gender and ethnic earnings gaps in Vietnam. Discrimination is questioned in the Vietnam's segmented labor market where women's participation in the informal sector is substantial and a majority of ethnic minority workers are engaged in non-waged occupations. However, data of earnings is mostly on wage employment and lack the informal sector; the previous studies on these topics have been limited in analyzing the gender and ethnic wage gaps and especially, the ethnic gap has been investigated without specific focus on disaggregating different ethnic groups. Taking advantage of the Labor Force Survey 2007, we try to overcome the mentioned issues to estimate the gender and ethnic earnings gaps and investigate further the heterogeneity in the gaps amongst different ethnic groups.

We first estimate the gap by using the traditional mincerian equation and then dealt with selection bias by using multinomial logit model to correct for self-selection into different labor market sectors. Without correcting for potential selection bias, ethnic minority people earns 22% less than Kinh while the gender gap is 16.4% and we find a great reduction in the ethnic gap when looking at non-farm job subsample, which is only 2.6%. It seems that the ethnic nonfarm labor income is marginal in Vietnam and this is confirmed by our multinomial logit model in which we find that the earning gap between Kinh and ethnic minority group is 48 percentage points in agriculture sector while it is 18 percentage points when they are engaged in formal and informal employment. Our results for agriculture and nonfarm occupation suggest that the remaining gap in farming and rural activities is probably more due to subtle differences in access to markets and land quality than to pure discrimination, whether by the employers or self-driven. In all models, the earning gap coefficients are extremely heterogeneous. Hmong, being the poorest and less assimilated ethnic minority in Vietnam (WB, 2009), have faced the largest gap in earning. The heterogeneity in ethnic earnings gaps raises the need of different ways/policy strategies to tackle the issue, which should be more specific for each ethnic group, rather than a "one size fits all" policy for all ethnic minority groups.

Decomposition results reflect different patterns between gender and ethnic earnings gaps. Using a multinomial logit model to correct for selectivity into different employment sectors, we find that the within sector differences explain most of the gender earning gaps and between sector differences favor men workers in terms of characteristics. The gender earnings gap is not large but the high proportion of unexplained gap is worth noticing, especially in the informal sector. When being engaged in the informal sector, the gap increases for wage earners and around 87% of the gap is left unexplained. Thus, our finding supports the common view that gender earnings gaps are likely to be found due to employer-based discrimination. On the other hand, ethnic earnings gap is explained almost equally by the within and between sector differences. Thus, it is seen that sectoral allocation is a concerning issue for the ethnic earnings gap and it corresponds the fact that ethnic earnings gaps are captured almost entirely in agriculture sector. Besides, not like the gender gaps, we find that differences in characteristics make up about two third of the earnings gap between Kinh and non-Kinh. These findings confirm the existence of ethnic earnings gap in Vietnam (Pham et al., 2007b) and the necessity to strengthen policies to increase education level and opportunities to access the non-farm jobs for ethnic minority groups as those factors are some of the roots for low returns to endowments of ethnic minority people (Baulch and Vu, 2011). Our findings are, to some extent, similar to what have been found in Latin American countries (Atal et al., 2009) that ethnic earnings gap is higher than gender earnings gap and occupational segregation plays an important role while it is not the case in some African countries (Nordman et al., 2011).

References

- Appleton, S., Hoddinott, J., Krishnan, P., (1999), “The gender wage gap in three African countries”, *Economic Development and Cultural Change* 47 (2), 289–312.
- ADB (Asian Development Bank) (2004), “The Factor Markets in Vietnam”, Discussion Paper No. 5, Making Markets Work Better for the Poor, Asia Development Bank, Vietnam Resident Mission, Hanoi.
- Atal, J.P., H. Ñopo and N. Winder (2009), “New Century, Old Disparities. Gender and Ethnic Wage Gaps in Latin America”, Research Department Working Paper 109, Washington, DC, United States: Inter-American Development Bank.
- Baulch, B., Truong, C., Haughton, D. and Haughton, J. (2004), “Ethnic minority development in Vietnam: a socio-economic perspective”, Chapter 8 in Glewwe, P., Agrawal, N., and Dollar, D. (eds), *Economic Growth, Poverty and Household Welfare in Vietnam*, World Bank Regional and Sectoral Studies, Washington DC, The World Bank.
- Baulch, B., Truong, C., Haughton, J. and Haughton, D. (2007), “Ethnic minority development in Vietnam”, *Journal of Development Studies*, 43(7): 1151-1171.
- Baulch, B., Pham, Hung T and Reilly, Barry (2012), “Decomposing the ethnic gap in rural Vietnam, 1993–2004”, *Oxford Development Studies*, 40 (1). pp. 87-117.
- Baulch B., Nguyen Thi Minh Hoa, Nguyen Thi Thu Phuong, and Hung Thai Pham, Reilly B. (2009), “Ethnic minority poverty in Vietnam”, processed, Hanoi, May.
- Baulch, B., Vu, D. H. (2011), “Poverty dynamics in Vietnam, 2002 – 2006”, in Baulch, B. (ed) (2011), “Why poverty persists: poverty dynamics in Asia and Africa”, Gloucester: Edward Elgar.
- Becker, G.S., (1957), “The Economics of Discrimination”, *University of Chicago Press*, Chicago.
- Blau F. and L. Kahn, (2001), “Understanding International Difference in the Gender Pay Gap”, *National Bureau of Economic Research*, Working paper 8200.
- Bourguignon, F., Fournier, M. and M. Gurgand, M. (2007), “Selection Bias Correction based on the Multinomial Logit Model: Monte Carlo Comparisons”, *Journal of Economic Surveys* (21).
- Cling J.-P., Marouani M.A, Razafindrakoto M., Robilliard A.-S. and Roubaud F. (2009), “The Distributive impact of Vietnam’s accession to the WTO”, *International Economy* (forthcoming).
- Friedman, J., D. Goodkind, T. C. Bui, and S. A. Truong. (2001), “Work and Retirement among the Elderly in Vietnam”, *Research on Aging*, 23(2):209-232.
- Gallup, J.L. (2004), “The Wage Labor Market and Inequality in Vietnam”, in Glewwe, P., Agrawal, N. and Dollar, D. (Ed.), *Economic Growth, Poverty, and Household Welfare in Vietnam*, *World Bank Regional and Sectoral Studies*, World Bank, Washington DC, pp. 53-94.
- GSO (2008), “Report on Labour Force and Employment Survey”, *Statistical Publishing House*, Hanoi.
- Heckman, J. (1979), “Sample Selection Bias as a Specification Error”, *Econometrica*, Vol. 47, pp.153-161.
- Hoang, H., Pham, G., Tran, M., and Hansen, H. (2007), ‘Ethnicity and poverty reduction’ in Nguyen, T. and Hansen, H. (eds) *Markets, Policy and Poverty Reduction in Vietnam*, Hanoi: Vietnam Academy of Social Sciences/National Political Publishing House.
- Liu, A.Y.C. (2004), “Gender Pay Gap in Vietnam: 1993-1998”, *Journal of Comparative Economics*,

Vol. 32 No. 3, pp. 586-596.

- McElwee P. (2004), “Becoming Socialist or Becoming Kinh”, in Duncan C., *Civilizing the Margins*, Ithaca, Cornell University Press.
- Mincer J. (1974), “Schooling, experience and earnings”, *Columbia University Press*, NY.
- Neumark, D., (1988), “Employers' discriminatory behavior and the estimation of wage discrimination”, *Journal of Human Resources* 23, 279–295.
- Nopo, H., Daza, N., Ramos, J., (2011), “Gender Earnings Gaps in the World”, IZA Discussion Papers 5736.
- Nordman, Robilliard A-S., Roubaud F. (2011), “Gender and Ethnic Earnings Gaps in Seven West African Cities”, *Labour Economics*, 18, Supplement 1, pp. S132-S145.
- Pham, T.H. and Reilly, R. (2007a), “Gender Pay Gap in Vietnam in 1993-2002: A Quantile Regression Approach”, *Journal of Asian Economics*, Vol. 18, No.5, pp.775-806.
- Pham, T.H. and Reilly, R. (2007b), “Ethnic Wage Inequality in Vietnam: Empirical Evidence from 2002”, Munich Personal RePEc Archive.
- Phung, D.T. and Do, T.T. (2014), “54 ethnic groups: Why different?”, Mekong Development Research Institute, Hanoi.
- Razafindrakoto M., Roubaud F., Nguyễn Hữu Chí (2011), “Vietnam Labor Market: An Informal Sector Perspective”, in Nguyễn Đức Thành (ed.), *Vietnam Annual Economic Report 2011: The Economy at a Crossroad*, Chapter 8, pp. 223-258, Edition Tri Thức, Hà Nội.
- Taylor, P. (2004), “Introduction: social inequality in a socialist state” in P. Taylor (ed): *Social inequality in Vietnam and the challenges to reform*, Institute of Southeast Asia Studies, Singapore.
- Tzannatos, Z. (1999), “Women And Labor Market Changes In The Global Economy: Growth Helps, Inequalities Hurt and Public Policy Matters”, Elsevier Science, the World Bank.
- Van de Walle, D. and Gunewardena, D. (2001), “Sources of ethnic inequality in Viet Nam”, *Journal of Development Economics*, 65: 177-207.
- Weichselbaumer, D., R. Winter-Ebmer and M. Zweimüller. (2007), “Market Orientation and Gender Wage Gaps: An International Study”, Department of Economics Johannes Kepler, University of Linz.
- World Bank (2009), “Country Social Analysis: Ethnicity and Development in Vietnam”, Washington DC, the World Bank.
- World Bank (2012), “Vietnam poverty assessment: well begun, not yet done - Vietnam's remarkable progress on poverty reduction and the emerging challenges”, Washington DC, the World Bank.

List of tables

Table 1: Sample size by ethnic groups in the LFS2007

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

Source: LFS2007, GSO; authors' calculations.

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

	Working age population	Employment rate	Under-employment rate	Second job	Public sector	Formal sector	Informal sector	Agriculture
By gender								
Males	72.9	77.13	11	17.2	11.2	16.6	23.8	48.4
Females	76.1	69.31	10.7	19.2	9.9	14.7	23.3	52.2
By ethnicity								
Kinh	75.5	71.2	10.9	17.4	11.5	17.5	26.4	44.6
Non-Kinh	68.8	82.1	10.5	23.1	4.7	5.3	6.5	83.5
Hoa	79.8	62.0	5.6	3.1	6.8	46.2	31.2	15.8
Thái	69.4	86.7	11.6	25.0	4.11	1.5	2.4	92.0
Tày	74.5	80.3	9.8	32.3	8.9	3.6	4.8	82.7
Mường	73.2	84.2	10.7	14.5	5.0	2.1	4.7	88.1
Khơ-me	72.2	76.2	16.9	16.7	3.5	7.3	18.3	70.9

Nùng	68.8	85.2	9.3	35.9	4.8	2.2	3.8	89.2
Hmông	55.0	92.3	10.7	35.8	0.9	0.1	0.4	98.5
Dao	64.6	88.0	3.0	28.5	2.2	0.5	2.1	95.2
Ê Đê	63.9	79.7	10.9	14.2	2.4	1.7	1.2	94.8
Gia-rai	59.1	90.8	10.5	21.6	2.9	0.0	0.9	96.3
Ba na	59.6	90.9	10.6	15.9	1.8	0.8	0.8	96.7
Xơ-ñăng	56.0	91.2	26.8	37.3	1.9	0.8	0.0	97.3
Ra-glai	56.9	88.1	6.4	20.9	1.3	0.3	2.7	95.8
Cơ-ho	60.6	84.6	12.1	8.9	1.6	0.3	0.9	97.2
Chăm	64.6	81.5	9.0	7.9	3.9	3.0	9.3	83.8
Giáy	63.3	86.3	0.2	13.5	3.0	1.6	3.1	92.3
Other Northern minorities	65.7	86.5	11.1	24.0	4.3	2.2	7.6	85.9
Other Central Highland	64.0	79.1	12.3	22.0	5.7	0.7	1.6	91.9
Other	63.5	87.7	1.6	7.7	3.9	1.6	2.5	92.1
Total	74.6	73.0	10.8	18.2	10.5	15.7	23.5	50.3

Source: LFS2007, GSO; authors' calculations.

Table 3: Demographic characteristics of the Labour Force by gender and ethnicity in 2007

	Rural area (%)	Age	Year of schooling	Married (%)	Household size
By gender					
Males	74.8	37.8	8.4	75.4	4.6
Females	75.4	38.3	7.8	73.6	4.5
By ethnicity					
Kinh	72.6	38.5	8.5	74.8	4.4
Non-Kinh	89.5	35.1	5.8	72.5	5.3
Hoa	21.9	38.5	7.5	57.0	4.4
Thái	97.0	34.3	5.8	75.8	5.3
Tày	91.3	35.7	8.1	74.3	5.5
Mường	97.3	35.6	8.0	72.8	4.6
Khơ-me	88.5	36.5	4.1	68.0	4.9
Nùng	93.5	35.0	6.2	74.9	4.8
Hmông	98.6	32.9	2.0	77.5	5.0
Dao	97.3	33.0	3.8	79.0	6.6
Ê Đê	86.8	34.7	3.9	70.1	5.7

Gia-rai	94.1	34.7	3.6	69.6	6.1
Ba na	93.5	33.9	3.2	73.6	5.8
Xơ-ñăng	96.6	34.9	4.1	69.9	5.5
Ra-glai	96.3	33.0	2.5	67.5	5.6
Cơ-ho	99.6	33.8	4.1	69.9	6.1
Chăm	89.5	35.0	4.5	72.5	6.4
Giáy	95.3	32.5	4.8	72.2	5.4
Other Northern minorities	96.1	33.8	5.8	74.9	5.3
Other Central Highland	93.4	34.6	4.1	70.7	5.6
Other	95.8	35.3	6.3	73.3	5.5
Total	75.1	38.0	8.1	74.5	4.5

Source: LFS2007, GSO; authors' calculations.

Table 4: Labour earnings models
(Log of hourly income; OLS)

	Model 1		Model 2		Model 3	
	Coef.	SE	Coef.	SE	Coef.	SE
Female	-0.191***	0.004	-0.167***	0.004	-0.179***	0.004
Years of schooling	-0.081***	0.002	-0.051***	0.002	-0.053***	0.002
(Years of schooling) ²	0.008***	0.000	0.007***	0.000	0.005***	0.000
Potential experience	0.048***	0.000	0.046***	0.000	0.049***	0.000
(Potential experience) ²	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
Institutional sector						
Foreign enterprise					0.024*	0.014
Domestic enterprise					-0.066***	0.009
Non-farm formal HB					-0.227***	0.007
Informal sector					-0.399***	0.007
Agriculture					-0.820***	0.007
Rural					0.041***	0.004
Intercept	0.863***	0.011	0.704***	0.011	1.231***	0.013
Hoa (Hán)	0.118***	0.019	-0.016	0.019	-0.086***	0.018
Thái	-0.375***	0.016	-0.483***	0.018	-0.291***	0.018
Tày	-0.500***	0.011	-0.420***	0.014	-0.266***	0.013
Mường	-0.548***	0.019	-0.514***	0.022	-0.312***	0.021
Khơ-me	0.078***	0.019	-0.051***	0.019	0.013	0.019
Nùng	-0.590***	0.017	-0.593***	0.019	-0.423***	0.019
Hmông	-1.415***	0.023	-1.153***	0.024	-0.949***	0.023
Dao	-0.923***	0.024	-0.808***	0.024	-0.602***	0.024
Ê Đê	0.276***	0.038	-0.127***	0.039	-0.003	0.038
Gia-rai	-0.196***	0.037	-0.471***	0.037	-0.270***	0.036
Ba na	0.035	0.039	-0.223***	0.039	-0.046	0.038
Xơ-ñăng	-0.460***	0.037	-0.639***	0.037	-0.466***	0.036
Ra-glai	-0.783***	0.045	-0.692***	0.044	-0.496***	0.043
Cơ-ho	0.270***	0.049	0.144***	0.048	0.277***	0.047
Chăm	0.004	0.049	-0.028	0.048	0.108**	0.046
Giáy	-0.676***	0.044	-0.634***	0.043	-0.450***	0.042

Other Northern minorities	-0.727***	0.033	-0.589***	0.032	-0.446***	0.031
Other Central Highland	-0.583***	0.028	-0.661***	0.028	-0.515***	0.027
Other	-0.192***	0.065	-0.161***	0.063	-0.001	0.061
Non Kinh	-0.432***	0.006	-0.387***	0.007	-0.254***	0.007
Province fixed effect	No		Yes		Yes	
No. Observation	310,154		310,154		310,154	
Adjusted R²	0.142		0.217		0.260	

Source: LFS2007, GSO; authors' calculations.

Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. Other coefficients for Kinh/non Kinh equation not reported.

Table 5: Non farm labour earnings models
(log of hourly income; OLS)

	Model 1		Model 2		Model 3 (only wage earners)	
	Coef.	SE	Coef.	SE	Coef.	SE
Female	-0.171***	0.004	-0.170***	0.004	-0.111***	0.003
Years of schooling	-0.055***	0.002	-0.046***	0.002	-0.017***	0.001
(Years of schooling) ²	0.006***	0.000	0.006***	0.000	0.004***	0.000
Potential experience	0.041***	0.000	0.039***	0.000	0.036***	0.000
(Potential experience) ²	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
Intercept	1.110***	0.012	1.083***	0.013	1.080***	0.009
Hoa (Hán)	-0.008	0.017	-0.049***	0.017	-0.097***	0.015
Thái	0.147***	0.032	-0.003	0.033	0.029	0.021
Tày	0.123***	0.014	0.016	0.017	0.053***	0.012
Mường	-0.025	0.032	-0.025	0.034	-0.082***	0.021
Khơ-me	-0.034	0.024	-0.057**	0.024	-0.053***	0.014
Nùng	0.123***	0.025	-0.041	0.027	-0.025	0.020
Hmông	-0.106	0.105	-0.273***	0.103	-0.039	0.062
Dao	0.109*	0.064	-0.013	0.063	-0.061	0.040
Ê Đê	0.014	0.096	-0.183*	0.095	0.121***	0.040
Gia-rai	0.117	0.118	-0.128	0.116	-0.080**	0.042
Ba na	-0.004	0.110	-0.229**	0.108	-0.129**	0.052
Xơ-năng	0.250	0.158	0.071	0.154	0.000	0.070
Ra-glai	-0.423***	0.151	-0.360**	0.147	0.075	0.069
Cơ-ho	-0.176	0.213	-0.154	0.208	-0.115	0.078
Chăm	-0.170**	0.080	-0.149*	0.079	0.008	0.045
Giáy	0.189**	0.087	0.028	0.086	-0.050	0.058
Other Northern minorities	0.037	0.055	-0.030	0.054	-0.031	0.034
Other Central Highland	0.106	0.075	-0.007	0.073	-0.126***	0.033
Other	0.253**	0.128	0.204	0.125	0.284***	0.071
Non Kinh	-0.060***	0.008	-0.031***	0.009	-0.026***	0.006
Province fixed effect	No		Yes		Yes	
No. Observation	199,014		199,014		117,253	
Adjusted R²	0.125		0.217		0.366	

Source: LFS2007, GSO; authors' calculations.

Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. Other coefficients for Kinh/non Kinh equation not reported.

Table 6a: Determinants of the allocation of people to different employment sectors
(Marginal effect from multinomial logit model on sectoral choice)

VARIABLES	(1) Public	(2) Formal	(3) Informal	(4) Agriculture	(5) Inactive
Female	0.0105*** (0.000903)	0.00950*** (0.00198)	-0.000310 (0.00292)	-0.0599*** (0.00316)	0.0402*** (0.00312)
Married	0.0268*** (0.00103)	0.0533*** (0.00225)	0.0585*** (0.00312)	0.0873*** (0.00325)	-0.226*** (0.00393)
Female*Married	-0.0151*** (0.00111)	-0.0627*** (0.00264)	-0.0444*** (0.00372)	-0.0111*** (0.00396)	0.133*** (0.00464)
Ethnic minority	0.00890*** (0.000989)	-0.0493*** (0.00271)	-0.115*** (0.00359)	0.156*** (0.00272)	-0.00114 (0.00424)
Education	0.00583*** (0.000384)	-0.00577*** (0.000657)	0.0103*** (0.000859)	-0.00651*** (0.000823)	-0.00384*** (0.00109)
Education ²	0.000579*** (2.30e-05)	0.00100*** (3.56e-05)	-0.00175*** (5.22e-05)	-0.00144*** (5.23e-05)	0.00161*** (6.16e-05)
Experience	0.00519*** (7.89e-05)	0.0115*** (0.000153)	0.0176*** (0.000203)	0.0116*** (0.000203)	-0.0458*** (0.000258)
Experience ²	-0.000128*** (1.64e-06)	-0.000294*** (2.75e-06)	-0.000322*** (3.26e-06)	-0.000179*** (2.90e-06)	0.000923*** (4.07e-06)
Rural area	-0.0118*** (0.000607)	-0.0765*** (0.00139)	-0.0424*** (0.00170)	0.292*** (0.00166)	-0.161*** (0.00233)
Spouse of the Head	-0.00756*** (0.000688)	-0.0201*** (0.00184)	-0.0297*** (0.00235)	0.00126 (0.00236)	0.0561*** (0.00349)
Children of the Head	-0.00118 (0.000927)	-0.00321 (0.00224)	-0.000585 (0.00303)	-0.0912*** (0.00329)	0.0962*** (0.00447)
Parents of the Head	-0.0133* (0.00681)	-0.0301* (0.0159)	0.0113 (0.0121)	-0.183*** (0.00977)	0.215*** (0.0127)
Family ties	0.00210** (0.00106)	0.00496* (0.00259)	0.00439 (0.00352)	-0.0777*** (0.00388)	0.0662*** (0.00501)
Non-family ties	-0.00100 (0.00554)	0.115*** (0.00965)	0.129*** (0.0161)	-0.0920*** (0.0244)	-0.151*** (0.0238)
HH size	-0.00139*** (0.000174)	0.00236*** (0.000379)	-0.00689*** (0.000504)	-0.00158*** (0.000500)	0.00750*** (0.000657)
Dependence ratio	-0.0164*** (0.00147)	-0.0433*** (0.00361)	0.130*** (0.00467)	0.0662*** (0.00456)	-0.136*** (0.00669)
Observations	453667	453667	453667	453667	453667

Source: LFS2007, GSO; authors' calculations.

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

Table 6b: Determinants of the allocation of people to different employment sectors
(Marginal effect from multinomial logit model on sectoral choice, disaggregated ethnic groups)

VARIABLES	(1) Public	(2) Formal	(3) Informal	(4) Agriculture	(5) Inactive
Female	0.0108* (0.00566)	0.00945* (0.00525)	-3.41e-05 (0.00283)	-0.0614* (0.0320)	0.0412* (0.0215)
Married	0.0275 (0.0202)	0.0539 (0.0472)	0.0582 (0.0817)	0.0885 (0.0753)	-0.228*** (0.0610)
Female*Married	-0.0157 (0.0127)	-0.0621 (0.0468)	-0.0442 (0.0621)	-0.0114 (0.0285)	0.133*** (0.0259)
Hoa	-0.0155* (0.00856)	0.0730* (0.0378)	-0.000636 (0.00669)	-0.0928* (0.0492)	0.0359* (0.0202)
Thai	0.0152 (0.0123)	-0.0915 (0.112)	-0.194 (0.273)	0.215*** (0.0153)	0.0558 (0.163)
Tay	0.0158*** (0.00588)	-0.105 (0.0992)	-0.135 (0.190)	0.183*** (0.0271)	0.0412 (0.112)
Muong	0.0117 (0.00858)	-0.0893 (0.0922)	-0.138 (0.194)	0.185*** (0.0275)	0.0306 (0.120)
Kho me	0.00492 (0.00369)	-0.0178 (0.0186)	-0.0244 (0.0351)	0.109** (0.0448)	-0.0721 (0.0623)
Nung	0.00982* (0.00528)	-0.106 (0.0865)	-0.0943 (0.133)	0.191*** (0.0520)	-0.000618 (0.0940)
Hmong	0.0408** (0.0177)	-0.225 (0.238)	-0.359 (0.507)	0.374*** (0.0219)	0.169 (0.269)
Dao	0.0160 (0.0118)	-0.193 (0.162)	-0.182 (0.257)	0.288*** (0.0582)	0.0716 (0.144)
Ede	0.0245 (0.0191)	-0.107 (0.153)	-0.289 (0.407)	0.230*** (0.0312)	0.142 (0.213)
Gia Rai	0.0758*** (0.0142)	-0.417 (0.332)	-0.287 (0.406)	0.450*** (0.0939)	0.178 (0.201)
Ba Na	0.0193 (0.0164)	-0.105 (0.131)	-0.222 (0.314)	0.290*** (0.0420)	0.0172 (0.213)
Xo Nang	0.197 (9.595)	0.262 (30.80)	-3.091 (180.0)	0.990 (47.42)	1.643 (92.22)
Ra glai	0.0445*** (0.0139)	-0.267 (0.227)	-0.236 (0.335)	0.320*** (0.0523)	0.139 (0.168)
Co-ho	0.0495*** (0.0176)	-0.338 (0.290)	-0.275 (0.392)	0.233*** (0.0408)	0.331*** (0.123)
Cham	0.00661 (0.0115)	-0.132 (0.105)	-0.101 (0.144)	0.172*** (0.0423)	0.0542 (0.0778)
Giay	-0.00284 (0.0130)	-0.115 (0.0912)	-0.0832 (0.121)	0.184*** (0.0560)	0.0168 (0.0811)
Other Northern	0.0222* (0.0114)	-0.108* (0.0638)	-0.0133 (0.0268)	0.146** (0.0701)	-0.0466 (0.0456)
Other Central Highland	0.0588*** (0.00995)	-0.253 (0.258)	-0.376 (0.530)	0.293*** (0.0417)	0.278 (0.229)
Other	0.0468** (0.0189)	-0.0705 (0.170)	-0.391 (0.553)	0.269*** (0.0649)	0.146 (0.314)
Control variables	Yes	Yes	Yes	Yes	Yes

Province fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	453667	453667	453667	453667	453667

Source: LFS2007, GSO; authors' calculations.

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

Table 7a: Earning equation estimates for different employment sectors

VARIABLES	Public	Formal sector	Informal sector	Agriculture
Female	-0.0805*** (0.0222)	-0.111*** (0.0155)	-0.204*** (0.0222)	0.0211 (0.0795)
married	-0.105*** (0.0231)	-0.117*** (0.0101)	-0.0340 (0.0266)	-0.161* (0.0869)
Female*Married	0.137*** (0.0264)	-0.0814*** (0.0115)	0.00687 (0.0268)	0.0589 (0.0918)
Ethnic minority	0.0209* (0.0121)	-0.206*** (0.00410)	-0.201*** (0.0262)	-0.658*** (0.0122)
Education	-0.0759*** (0.0105)	0.0286*** (0.00267)	0.0369*** (0.00517)	0.0527*** (0.000779)
Education2	0.00375*** (0.000574)	0.000297 (0.000182)	-0.00308*** (0.000390)	-0.000615 (0.000486)
Experience	0.0229*** (0.00240)	-0.0204*** (0.00553)	0.0127*** (0.00110)	-0.0131 (0.00988)
Experience 2	-0.000243*** (5.63e-05)	0.000553*** (0.000133)	-0.000166*** (1.75e-05)	0.000530*** (9.77e-05)
Rural area	0.0564*** (0.00851)	-0.119*** (0.0364)	-0.110*** (0.0381)	-0.763*** (0.0484)
BFG_public	-0.160*** (0.0170)	-1.188*** (0.0747)	-0.881*** (0.0761)	-0.605*** (0.0764)
BFG_Formal	-0.355*** (0.0911)	0.0110 (0.0335)	0.173* (0.105)	0.434* (0.226)
BFG_Informal	0.397*** (0.0698)	-0.641*** (0.0638)	0.0345 (0.0222)	1.264*** (0.167)
BFG_Agriculture	0.177*** (0.0297)	-0.932*** (0.0633)	-0.355*** (0.123)	-0.224*** (0.0289)
BFG_Inactive	0.234*** (0.0604)	0.563*** (0.138)	0.580*** (0.0379)	2.180*** (0.0712)
Constant	2.493*** (0.228)	1.196*** (0.167)	1.250*** (0.106)	2.643*** (0.539)

Source: LFS2007, GSO; authors' calculations.

**Table 7b: Earning equation estimates for different employment sectors
(disaggregated ethnic groups)**

VARIABLES	Public	Formal sector	Informal sector	Agriculture
Female	-0.0816*** (0.0104)	-0.110*** (0.0119)	-0.203*** (0.0128)	0.0945 (0.280)
Married	-0.106*** (0.0116)	-0.120*** (0.0142)	-0.0334*** (0.0109)	-0.181 (0.189)
Female*Married	0.138*** (0.00825)	-0.0791*** (0.0153)	0.00813 (0.0241)	0.0563 (0.281)
Hoa	-0.00698	0.111*	0.0374	0.0870

	(0.0536)	(0.0567)	(0.0358)	(0.251)
Thai	0.0622***	-0.216***	-0.269***	-0.662***
	(0.0189)	(0.00752)	(0.0433)	(0.0758)
Tay	0.0781**	-0.329***	-0.369***	-0.525***
	(0.0322)	(0.0619)	(0.0236)	(0.0418)
Muong	-0.107***	-0.214***	-0.139***	-0.709***
	(0.0278)	(0.0415)	(0.0139)	(0.0785)
Kho me	-0.0617*	-0.210***	-0.107***	-0.423***
	(0.0340)	(0.0735)	(0.00796)	(0.0471)
Nung	0.0771	-0.298***	-0.258***	-0.708***
	(0.0622)	(0.0672)	(0.0492)	(0.0789)
Hmong	-0.00744	-1.141*	-0.783*	-1.316***
	(0.0771)	(0.605)	(0.473)	(0.125)
Dao	0.0805	-0.215***	-0.196*	-0.900***
	(0.0632)	(0.0372)	(0.107)	(0.0774)
Ede	-0.188**	-0.118*	-0.495***	-0.606***
	(0.0794)	(0.0637)	(0.0332)	(0.118)
Gia Rai	-0.0857	-0.786**	-0.531***	-1.167***
	(0.124)	(0.376)	(0.0693)	(0.116)
Ba Na	-0.179	-0.790	-0.121*	-0.674***
	(0.256)	(0.487)	(0.0718)	(0.0539)
Xo Nang	0.0426	-0.0452	0	-0.938***
	(0.0847)	(0.0568)	(0)	(0.174)
Ra glai	-0.191**	-0.0207	-0.667***	-1.187***
	(0.0905)	(0.131)	(0.193)	(0.0102)
Co-ho	0.0561	-0.584***	-0.493***	-0.0620
	(0.0698)	(0.155)	(0.143)	(0.120)
Cham	0.104	-0.196	-0.342***	-0.260***
	(0.221)	(0.136)	(0.0501)	(0.0259)
Giay	0.201***	0.0391	-0.0833	-0.872***
	(0.0766)	(0.141)	(0.0831)	(0.129)
Other Northern	0.0257	-0.352***	-0.165***	-0.750***
	(0.0843)	(0.116)	(0.0428)	(0.0733)
Other Central Highland	0.0424	-0.347	-0.825***	-0.793***
	(0.0451)	(0.305)	(0.257)	(0.0826)
Other	0.247**	-0.229***	-0.385	-0.459***
	(0.124)	(0.0779)	(0.268)	(0.0935)
Education	-0.0772***	0.0281***	0.0354***	0.00915
	(0.00425)	(0.00526)	(0.00314)	(0.00916)
Education2	0.00380***	0.000271	-0.00298***	0.00293***
	(0.000279)	(0.000407)	(0.000381)	(0.000673)
Experience	0.0230***	-0.0212***	0.0122***	-0.0194*
	(0.000759)	(0.00458)	(0.00231)	(0.0109)
Experience2	-0.000244***	0.000562***	-0.000159***	0.000667***
	(2.54e-05)	(0.000113)	(4.58e-05)	(0.000183)

Rural area	0.0588*** (0.00937)	-0.113*** (0.0411)	-0.104*** (0.0156)	-0.618*** (0.102)
BFG_public	-0.159*** (0.0140)	-1.199*** (0.0952)	-0.866*** (0.0544)	-0.150 (0.119)
BFG_Formal	-0.367*** (0.118)	0.0130 (0.0450)	0.198*** (0.0688)	-0.207*** (0.0631)
BFG_Informal	0.396*** (0.0158)	-0.641*** (0.0305)	0.0337*** (0.000414)	0.716*** (0.102)
BFG_Agriculture	0.185*** (0.0244)	-0.925*** (0.0871)	-0.331*** (0.0480)	-0.195*** (0.0620)
BFG_Inactive	0.234*** (0.0297)	0.589*** (0.111)	0.611*** (0.0221)	2.701*** (0.322)
Constant	2.495*** (0.114)	1.221*** (0.203)	1.292*** (0.0409)	1.974*** (0.698)

Source: LFS2007, GSO; author's calculations.

Table 8 Labour earning quantile regression model

	10th	25th	50th	75th	90th
Female	-0.0999*** (0.00424)	-0.132*** (0.00288)	-0.143*** (0.00248)	-0.138*** (0.00283)	-0.128*** (0.00383)
Years of schooling	-0.0431*** (0.00223)	-0.0343*** (0.00145)	-0.0242*** (0.00121)	-0.00891*** (0.00135)	0.00681*** (0.00184)
(Years of schooling) ²	0.00449*** (0.000120)	0.00403*** (7.80e-05)	0.00352*** (6.54e-05)	0.00288*** (7.41e-05)	0.00221*** (0.000101)
Potential experience	0.00887*** (0.00103)	0.00913*** (0.000714)	0.00996*** (0.000618)	0.00910*** (0.000701)	0.00627*** (0.000936)
(Potential experience) ²	-0.000105*** (1.81e-05)	-0.000113*** (1.26e-05)	-0.000105*** (1.09e-05)	-4.78e-05*** (1.24e-05)	5.50e-05*** (1.66e-05)
Institutional sector					
Foreign enterprise	0.124*** (0.0163)	-0.0278** (0.0109)	-0.136*** (0.00939)	-0.191*** (0.0108)	-0.161*** (0.0149)
Domestic enterprise	-0.0181* (0.0100)	-0.0953*** (0.00664)	-0.151*** (0.00564)	-0.158*** (0.00643)	-0.103*** (0.00886)
Non farm formal HB	-0.193*** (0.00895)	-0.157*** (0.00579)	-0.120*** (0.00484)	-0.0490*** (0.00547)	0.0548*** (0.00745)
Informal sector	-0.352*** (0.00833)	-0.329*** (0.00528)	-0.322*** (0.00434)	-0.299*** (0.00488)	-0.245*** (0.00667)
Agriculture	-0.781*** (0.00877)	-0.752*** (0.00548)	-0.679*** (0.00451)	-0.570*** (0.00516)	-0.427*** (0.00718)
Rural	-0.0403*** (0.00499)	-0.0141*** (0.00339)	-0.00382 (0.00296)	0.000702 (0.00340)	-0.00694 (0.00460)
Intercept	1.298*** (0.0255)	1.624*** (0.0173)	1.915*** (0.0147)	2.152*** (0.0165)	2.401*** (0.0221)
Mills ratio	-0.423*** (0.0142)	-0.319*** (0.00986)	-0.279*** (0.00858)	-0.297*** (0.00979)	-0.346*** (0.0132)
Hoa (Hán)	-0.117*** (0.0201)	-0.0580*** (0.0137)	-0.0286** (0.0118)	0.0340** (0.0134)	0.0724*** (0.0181)
Thái	-0.192*** (0.0199)	-0.289*** (0.0135)	-0.311*** (0.0116)	-0.317*** (0.0134)	-0.292*** (0.0187)
Tây	-0.280*** (0.0142)	-0.254*** (0.00986)	-0.268*** (0.00873)	-0.220*** (0.0101)	-0.162*** (0.0135)

Mường	-0.284*** (0.0236)	-0.294*** (0.0161)	-0.353*** (0.0140)	-0.317*** (0.0164)	-0.268*** (0.0225)
Khơ-me	0.00614 (0.0207)	-0.00849 (0.0141)	-0.0232* (0.0121)	-0.0422*** (0.0137)	-0.0619*** (0.0186)
Nùng	-0.432*** (0.0202)	-0.449*** (0.0138)	-0.411*** (0.0122)	-0.275*** (0.0141)	-0.208*** (0.0191)
Hmông	-0.603*** (0.0254)	-0.778*** (0.0172)	-0.683*** (0.0151)	-0.534*** (0.0177)	-0.462*** (0.0246)
Dao	-0.526*** (0.0265)	-0.429*** (0.0179)	-0.449*** (0.0154)	-0.481*** (0.0177)	-0.471*** (0.0241)
Ê Đê	-0.179*** (0.0413)	0.0778*** (0.0283)	0.181*** (0.0245)	0.102*** (0.0277)	-0.0402 (0.0372)
Gia-rai	-0.270*** (0.0400)	-0.305*** (0.0270)	-0.274*** (0.0234)	-0.131*** (0.0265)	-0.116*** (0.0357)
Ba na	0.0520 (0.0417)	0.0515* (0.0285)	0.0517** (0.0247)	-0.0289 (0.0281)	-0.172*** (0.0379)
Xơ-ñăng	-0.220*** (0.0408)	-0.206*** (0.0274)	-0.345*** (0.0235)	-0.421*** (0.0267)	-0.518*** (0.0364)
Ra-glai	-0.461*** (0.0479)	-0.394*** (0.0321)	-0.339*** (0.0277)	-0.278*** (0.0319)	-0.195*** (0.0437)
Cơ-ho	0.148*** (0.0529)	0.0875** (0.0356)	0.266*** (0.0304)	0.285*** (0.0345)	0.513*** (0.0464)
Chăm	0.128** (0.0509)	0.0707** (0.0347)	0.0775*** (0.0300)	0.140*** (0.0341)	0.0749 (0.0457)
Giáy	-0.241*** (0.0468)	-0.332*** (0.0316)	-0.300*** (0.0274)	-0.383*** (0.0313)	-0.320*** (0.0427)
OtherNorthern	-0.388*** (0.0348)	-0.380*** (0.0236)	-0.323*** (0.0203)	-0.248*** (0.0233)	-0.230*** (0.0319)
Other Central Highland	-0.524*** (0.0292)	-0.344*** (0.0200)	-0.310*** (0.0174)	-0.292*** (0.0200)	-0.231*** (0.0273)
Other	0.00922 (0.0663)	-0.0153 (0.0459)	-0.0661* (0.0395)	-0.0693 (0.0449)	0.0296 (0.0600)
Non Kinh	-0.242*** (0.0102)	-0.231*** (0.00406)	-0.225*** (0.00419)	-0.197*** (0.00517)	-0.172*** (0.00945)
Province fixed effect	Yes				
No. Observation	310,154				

Source: LFS2007, GSO; authors' calculations.

Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. Other coefficients for Kinh/non Kinh equation not reported.

Table 9: Neumark decomposition of gender and ethnic earnings gaps

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

Source: LFS2007, GSO; authors' calculations.

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

Table 10: Neumark decomposition of ethnic earning gaps between different ethnic groups

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

Source: LFS2007, GSO; authors' calculations.

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

Table 11a: Neumark decomposition of gender earnings gaps and ethnic earnings gaps by sector

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

Source: LFS2007, GSO; authors' calculations.

Table 11b: Neumark decomposition of gender earnings gaps in the private and informal sectors

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

Source: LFS2007, GSO; authors' calculations.

Table 12: Full decomposition of gender earnings gap

	Without correcting for selectivity		Correcting for selectivity	
		%		%
Gender earnings gap				
=log(Male group earning) – log(Fem. group earning)	0.224	%	0.168	%
Difference due to within sector differences in earnings attributable to				
Characteristics	0.058	26.1	0.019	11.3
Deviation in male returns	0.078	34.8	0.069	41.4
Deviation in female returns	0.079	35.4	0.069	40.9
Sub-total	0.215	96.3	0.157	93.6
Difference due to differences between sectoral allocation				
Characteristics	0.017	7.8	0.049	29.4
Deviation in effect of characteristics on male location	-0.004	-1.7	-0.021	-12.5
Deviation in effect of characteristics on female location	-0.006	-2.5	-0.018	-10.6
Sub-total	0.007	3.6	0.01	6.3

Source: LFS2007, GSO; authors' calculations.

Table 13: Full decomposition of ethnic earnings gap

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

Source: LFS2007, GSO; authors' calculations.