Journal of Applied Economics and Business Research
JAEBR, 10(1): 7-18(2020)

Income Gap in Vietnam Using Bayesian Quantile Regression

Duong Quynh Nga (Dr.)[1]
Ho Chi Minh City Open University, Vietnam

Tran Hoang Truc Linh
Ho Chi Minh City Open University, Vietnam

Nguyen Tran Cam Linh
Ho Chi Minh City Open University, Vietnam

Đoan Truong An
Department of Statistics District 10, Vietnam

Abstract
This paper examines the wage gaps in informal – formal sector of Vietnam in 2014 by using Machado and Mata (2005) decomposition based on Bayesian quantile regression to investigate the sources of the gap at different quantiles. Our analysis uses data from Labor Force Survey of Vietnam in 2014 for assessing the magnitude of various formal/informal sector earnings gaps. Is there an informal sector job earning penalty? The result shows that workers in the informal economy are always more disadvantaged than workers in the formal economic sector. While at the highest wage percentile, the wage gap between employees in the two economic sectors is mainly due to the differences in employee characteristics (explain 85.56%) and at the cash low percentile, the only cause of wage differences is due to the level of remuneration in wage payments (differences due to regression coefficients). We also find that the largest part (78% averagely) of the recent changes in Vietnam income inequality is explained by the wage structure effect and second that the income inequality is rising more at the top end of the wage distribution. The decline in the unionization rate and industry factor has a small impact on total wage inequality while differences in returns to education, position, age and gender discrimination are the dominant factors accounting for these recent changes.

Keywords: formal sector, informal sector, Bayesian quantile regression, Machado and Mata decomposition.

JEL codes: D63, E01, E24, O17

1. Introduction

According to the international definition, the informal economic sector is understood here to include all production and business households without legal status, without business registration in non-agricultural and forestry sectors but still produce goods and services for sale or exchange in the market. According to the international understanding, the informal sector is that all production and business establishments do not carry out business registration. Agricultural

[1] Corresponding to Duong Quynh Nga, Email: nga.dq@ou.edu.vn

activities do not count on this area due to its peculiar nature. The formal economic sector consists of companies with legal status; non-profit organizations; administrative and non-business units of state agencies; and production and business establishments do not have legal status but have business registration at management agencies at all levels (usually individual production and business establishments).

The difference in manpower between workers in the formal and informal economic sectors is extremely obvious and is one of the reasons which makes the differences in wages for workers between the two groups. However, economists like Becker and Cain also argue that, apart from human capital (and / or labor productivity), the difference in wages is due to discrimination (Becker, 1971, Cain, 1986, quoted by Tran Thi Tuan Anh, 2015). Workers in the economic sector are officially guaranteed about the salary regime as well as benefited from social welfare regimes. From the perspective of the employees' benefits when signing a labor contract, the employees will be paid, increased wages, working properly with working conditions, working time, resting time, attendance social insurance in accordance with the agreements stated in the labor contract, the content of the labor contract has been recognized and protected by the law, in other words, the labor contract is basically legal department bind employer and employees in their relationship. Social insurance creates a lot of benefits for workers such as receiving social insurance or subsidies, which are calculated based on the time of social insurance when they reach retirement age. Maternity or quit working for other legitimate reasons, etc. In addition, when workers participate in formal economic sectors such as state administrative and non-productive units; political organizations, social organizations or corporations, state companies, private companies, etc., the role of federation is extremely important for workers. Trade unions supervise the signing of labor contracts between businesses and employees, helping participating workers to compromise their legitimate aspirations, including salary, bonuses, and so on. When the role of the Federation promotes effectively, the production movement of institutions is encouraged; after that, the establishments promote working efficiency, contributing to improve the regimes on wages for employees.

According to the data of the Labor Force Survey (LFS) of the General Statistics Office from 2011 to 2015 in Ho Chi Minh City, the wages of workers in business and production households (where there are up to two thirds of the informal economy) are always lower than the wages of employees in state-owned enterprises, private, collective as well as foreign invested area. Thus, there is an existence of wage gaps between formal and informal economic workers. However, according to economists like Becker (1971), Cain (1986) distinguishes two ways of explaining the problem of wage disparity: the difference in wages due to human capital (or labor productivity) and wage differences due to discrimination.

The informal sector in Vietnam helps the poorest people to earn a living, especially in urban areas. However, working condition in this area is very precarious, the income is not only low but [2]also unstable. Because they are not recognized, registered or managed, and be not protected by the labor market institutions, workers in the informal sector have to face the risk of becoming poor working class. Employees in the informal sector and their families often suffer disadvantages cause they are not regulated by labor laws. Besides, they do not receive any support from the social assistance programs. In addition, the voice of workers in the informal sector is rarely mentioned in the decision-making process. Enterprises in the economy are officially faced with unfair competition from the non-main sector due to the price cuts of goods and services because they do not contribute to social insurance and taxes. The life of workers in the informal sector is characterized mainly by low income and poor working conditions, making them more vulnerable to external impacts. Roxana Maurizio (2010) in the report presented at the Conference on Informal and Informal Employment in Vietnam in Hanoi also concluded that there is always a wage segment among workers in the market. Roxana Maurizio has argued that, when there is no wage segment and no barriers to labor migration, without participating in the formal economic sector, excess workers will move to the informal economy that has lower productivity, then it will cause a general reduction in wages of informal sector workers. Roxana Maurizio also argues that the typical informal sectors are small business establishments, individuals, households with low productivity and therefore they have lower average remuneration. In addition, the failure to perform business registration obligations, not fulfill tax obligations, etc. and of course the employees will not be protected by social security, union, and so on. All of those may reduce the efficiency and productivity of businesses and lead to lower wages for workers (Beccaria and Groisman, 2008, quoted by Roxana Maurizio, 2010).

The aim of our study was to decompose the observed inequalities in income distributions of worker in formal and informal sector applying the Machado-Mata technique. We employ data from Labor Force Survey of Vietnam in 2014.

The structure of the paper is as follows. Section 2 briefly discusses the literature of decomposing the wage gap throughout the wage distribution. Section 3 presents econonometric methodology. Section 4 presents data and the results of the decomposition of differences in income denstites between formal and informal sector. Section 5 concludes.

2. Literature Review

A number of studies about the differences in wage between formal and informal economic workers in countries such as Thailand, China, Korea, Turkey, and Latin American countries (Argentina, Brail, Chile, and Peru) concluded that there is always a wage gap between formal and informal economic workers. According to Dasgupta et al (2015), workers in the informal sector in Thailand receive monthly lower wages than formal workers, which are different due to their characteristics that is only explained by 67.9% the rest is considered unequal in pay. Zuo (2013), the Chinese case study concluded that the difference in hourly wages between formal and informal labor is mainly differentiated by the labor market segmentation effect (67%), the difference due to employee's characteristics is only explained by 33%. Cho and Cho (2009) conclude that there is a significant gap between wages between formal and informal economic workers in the Korean labor market. Tansel and Kan (2012), the study proves the existence of wage gap between workers in the formal and informal economic sections in Turkey: informal sector workers receive the salary is 31.8% lower than that of the formal economic area and more than 50% of this can be explained by the characteristics of the workers, the rest is unexplained due to the regression coefficient (also called the labor market segment). Murphy and Welch (1992) attribute the changes to the skill-biased technological progress which increases the rate of growth of the relative demand for highly educated and “more-skilled” workers (see also Mincer, 1993, Katz and Autor, 1999). Murphy and Welch (1992) stress the impact of the globalization which increases the rate of unskilled immigration workers and led to a decrease in the growth of the relative supply of skills (see also Katz and Murphy, 1992).

 ISSN 1927-033X

DiNardo et al. (1996) focus on changes in labour market institutions, in wage setting norms including the decline in unionization, on the erosion of the real and relative value of the minimum wage. According to research by Nguyen Thanh Binh (2014), there are differences in wages between workers in the formal and informal economic sectors in HCMC. Ho Chi Minh. Employees in the informal sector receive only about 70% of monthly salary compared to workers in the formal economic sector in 2012. Nguyen Huu Chi (2012), informal employment research for labor migration from rural to urban with the following results: for migrant workers from rural areas: the wage gap between workers in the formal and informal economic sectors is 16% and only 23% of the difference in distance is explained by the impact of individual ability; for migrant workers from urban areas: the wage gap between workers in the formal and informal economic sectors is 26% and only 22% of the difference in distance explained by the impact belongs to personal abilities. Maurizio (2010), research in 4 Latin American countries: Argentina, Brazil, Chile and Peru, Oaxaca-Blinder analysis results: there is a difference in income between workers in the formal economic sector and informal in 4 countries: Argentina, Peru, Brazil and Chile. The analysis results of "coefficient effect" are statistically significant and are evidence for the effect of income segmentation of 2 economic sectors of 4 countries. The "resource effect" has also been shown to be statistically significant and in these cases, this effect is the factor that explains the highest proportion of income disparities.

Hence, besides human capital factors, the difference in wages of workers in the formal and informal sector is also due to inequality in payment.

3. Decomposition of differences in distributions using quantile regression

Introduced by Koenker and Bassett (1978), quantile regression models aim at modeling the effect of the explanatory variables on the conditional distribution of the outcome variable. They have been increasingly used in empirical labour market studies, to describe parsimoniuously the entire wage conditional distribution (see e.g. Buchinsky, 1994, Chamberlain, 1994, Machado and Mata, 2001). Several competing methods of estimation in both classical and Bayesian frameworks have been recently developed (see for instance Yu and Moyeed, 2001, Kozumi and Kobayashi, 2011 for the Bayesian side). Since any quantile can be used in any part of the outcome distribution, the quantile regression models are more flexible and more robust to outliers than the classical mean regression models.

While the conditional quantile regression models can be useful, they are very restrictive. First a change in distribution of covariates may change the interpretation of the coefficient estimates. This point is illustrated for instance in Powell (2011). To overcome this restriction, Firpo et al. (209) have proposed a new regression method which evaluates the impact of changes in the distribution of the explanatory variables on the quantles of the unconditional distribution of the outcome variable. Second, the property that, in the popular Oaxaca-Blinder decomposition method of a simple linear regression, differences in unconditional means is equal to differences between conditional means, is no longer valid for conditional quantile regressions. As explained in e.g. Bazen 2011, with conditional quantile regressions, the difference in unconditional quantiles is not equal to difference in conditional quantiles. This question has received several answers in the literature, see e.g. Juhn et al. (1993), DiNardo et al. (1996), or Machado and Mata (2005), but none of these methods can be used to decompose general distributional measures in the same way that the means can be decompose general distributional measures in the same way that the means can be decomposed using the conventional Oaxaca-Blinder method. However, the method of Melly (2005) and the RIF-regression method of Firpo et al. (2009) can perform a detailed

 ISSN 1927-033X

decomposition very much in the spirit of the traditional Oaxaca decomposition for the mean (Firpo et al. 2011).

3.1. Conditional quantile regression estimation

One natural workaround for wage density decomposition that does not work is conditional quantile regression. Quantile regression methods were developed to answer questions about effects of covariate on other features of the outcome distribution aside from the means. As in OLS, quantile regression coefficients can be modeled as a linear approximation to the conditional quantile function. If we let $\tau \in (0,1)$ denote the τ^{th} quantile of the distribution of log wages given the vector of covariates, then the linear model is

$$Q_\tau = arg \min_b \mathbb{E}\{\rho_\tau(Y_i - X_i'b)\}$$

Where

$$\rho_\tau(Y_i - X_i'b) = \begin{cases} \tau(Y_i - X_i'b) & for\ (Y_i - X_i'b)\ \geq 0 \\ (1-\tau)(Y_i - X_i'b) & for\ (Y_i - X_i'b) < 0 \end{cases}$$

The function ρ_τ is referred to as the "check function" because the weights will be shaped like a check with the infection point at $Y_i - X_i'b = 0$ (more on this below). Note another key difference between the OLS and CQR minimands: the CQR minimand averages absolute deviations rather than squared deviations. That's why it delivers an estimate that's not sensitive to outliers. You don't reduce the minimand by moving $\hat{\beta}_\tau$ towards the extremes because you pay with deviations from the opposite end.

Please also be clear about the correct interpretation of β_τ. Let's say that X is a dummy variable indicating college versus high-school attendance and the outcome variable is earnings. You estimate a quantile regression for $\beta_{0.5}$. the correct interpretation of $\hat{\beta}_{0.5}$ is generally not that is it the effect of college attendance on the median wage earner. Rather, it is the effect of college attendance on the median of the wage distribution. The former interpretation requires an additional rank invariance assumption: that the 50th percentile college earner would also have been the 50th percentile high school earner, so that the treatment (college) preserve rank orderings. This is a very strong and generally untestable assumption.

Unlike the conditional expectation function, the conditional quantile function is not a linear operator. That means that $Q_\tau(Y_i|X_i) = X_i'\beta_\tau$ does not imply that $Q_\tau(Y_i) = Q_\tau(X_i')\beta_\tau$. Consequently, we can't use CQR to answer question like, "What is the effect of sending more women to college on the 10th percentile of the wage distribution?" instead, CQR estimates the effect of the college education on the 10th percentile of the wage distribution for a given set of covariates – e.g. young women. That's unfortunate because we typically care about how an economic variable affects the distribution of wages, not the distribution of wages conditional on X. thus if the objective is wage density decomposition the CQR is not the straightforward place to start.

Observed wage differences between subgroups can be considered as sum of explicable differences and pure discrimination, both unobserved. Thus, the key issue is to quantify the contribution of various explanatory wage-relevant characteristics to this sum. Only the part not explicable by different characteristics of the subgroups can be regarded as (pure) discrimination. To estimate the two components requires an "as if" calculation: what, for example, would the wage

 ISSN 1927-033X

distribution of women look like, if they received the same remunation for each characteristic as men? One might also pose the same question differently: what would the wage distribution of men look like, if they had equal schooling and experience etc. (i.e. characteristics) as women? The phrasing does not matter. The important thing to note is, that, in econometrics terms, this requires the estimation of counterfactual distributions.

Machado and Mata (2005) provide one possible solution to this problem they augment conditional quantile estimates for the coefficients β_k with corresponding unconditional densities (actual and counterfactual) derived from resampling. The MM-approach is much more widely used than alternatives without (conditional) quantile regressions, including the reweighting technique of DiNardo et al. (1996), or RIF-regressions with reweighting by Fortin et al. (2011). Like MM these alternative techniques have their own virtues and drawbacks due to the underlying particular assumptions. Overall, we considered the MM approach more intuitive than these other approaches. The MM-approaches can be summarized as follows:

Let us consider the relationship between the regressors and outcome using the conditional quantile function:

$$Q_\theta(y|X) = \Phi^{-1}_{y|X}(\theta, X) = X\beta(\theta)$$

Where $Q_\theta(y|X)$ – the θ^{th} quantile of a variable y conditional on covariates X, $\theta \in (0,1)$; $\Phi_{y|X}$ – the joint distribution function for the variable y.

A different quantile θ may be specified (most frequently from three to nine). For each quantile other parameters $\beta(\theta)$ are estimated. These coefficients can be interpreted as the returns to different characteristics X at given quantiles of the distribution of y. bootstrap standard errors are often used (Gould, 1992; 1997). Quantile regression is more robust than least squares regression to non-normal errors and outliers. This method also provides a richer characterization of data, considering the impact of covariates on the entire distribution of y, not only its conditional mean.

Machado and Mata (2005) used quantile regression in order to estimate counterfactual unconditional wage distributions. Since the unconditional quantile is not the same as the integral of the conditional quantiles, authors provide a simulation-based estimator where the counterfactual distribution is constructed from the generation of a random sample. This estimator is widely used in various applications (cf. Albrecht, Bjorklund and Vronman, 2003; Melly, 2005). The idea underlying this technique is the probability integral transformation theorem. If U is a uniform random variable on [0,1], then $F^{-1}(U)$ has distribution F. thus, if $\theta_1, \theta_2, \ldots, \theta_m$ are drawn from a uniform (0,1) distribution, the corresponding m, estimates of the conditional quantiles of wages at $X, \{X\hat{\beta}(\theta_i)\}, i = 1, \ldots., m$ constitute a random sample from the (estimated) conditional distribution of wages given X (Machado and Mata, 2005, p.448).

The Machado-Mata approach to generate a random sample from the wage density that would prevail in group A if model (7) was true and covariates were distributed as $h^A(X)$ is as follows:

Generate a random sample of size m from a U [0,1]: $u_1, \ldots, u_m$.

Using the dataset for group A, estimate m different quantile regression $Q_{ui}(y|X_A)$, obtaining coefficients $\widehat{\beta_A}(u_i), i = 1, \ldots, m$

 ISSN 1927-033X

Generate a random sample of size m with replacement from the rows of X_A, denoted by $\{X_{Ai}^*\}, i = 1, \dots, m$

$\{y_{Ai}^* \equiv X_{Ai}^* \widehat{\beta_A}(u_i)\}, i = 1, \dots, m$ is a random sample of size m from the unconditional distribution $f^A(y)$

Counterfactual distributions could be estimated by drawing X from another distribution and using different coefficient vectors. For example, to generate a random sample from the wage density that would prevail in group A and covariates that were distributed as $h^B(X)$ (e.g. the men log wage density that would arise if men were given women's labor market characteristics but continued to be paid like men), we generate a random sample from the rows of X_B, denoted by $\{X_{Ai}^*\}, i = 1, \dots, m$ and $\{y_{Ai}^* \equiv X_{Ai}^* \widehat{\beta_A}(u_i)\}$ is a random sample from the counterfactual distribution $f^c(y)$.

Consequently, the idea of Machado-Mata decomposition of the difference between the wage densities in two groups, $\hat{\Delta}^\theta = Q_\theta(y_A^*|X_A^*) - Q_\theta(y_B^*|X_B^*) = Q_\theta(y_A^*|X_A^*) - Q_\theta(y_A^{C*}|X_B^*) + Q_\theta(y_A^{C*}|X_B^*) - Q_\theta(y_B^{C*}|X_B^*) = \underbrace{(X_A^* - X_B^*)\widehat{\beta_A}(\theta)}_{\hat{\Delta}^*_{explained}} + \underbrace{\left(\widehat{\beta_A}(\theta) - \widehat{\beta_B}(\theta)\right) X_B^*}_{\hat{\Delta}^*_{unexplained}}$

We can also use the Machado-Mata approach to estimate standard errors for the estimated densities by repeating the procedure many times and generating a set of estimated densities. The standard error of the estimator diminishes as we increase the number of replications but estimating a large number of replications is time-consuming especially when the number of observations is high (Melly, 2006).

3.2. Bayesian Inference for conditional quantile regression

To make inferences on the parameter of interest β_τ and σ_τ, given τ and the observations on (X, Y), one has to specify a prior density for β. The posterior distribution of β_τ , $\pi(\beta|y)$ is proportional to

$$\pi(\beta, \sigma|y) \propto L(y|\beta)\pi(\beta|\sigma)$$

Where $L(y|\beta)$ is the likelihood function given in (6) and $\pi(\beta|\sigma)$ is the prior distribution of β and. Yu and Moyeed (2001) show that for any type of prior, including an improper prior, the posterior moments exist. They choose an improper prior and no conjugate prior is available when the model is presented in this form. The posterior density has to be integrated out by a MCMC method. Yu and MOyeed (2001) make use of the simple random walk Metropolis with a Gaussian proposal. The method is available as package bayesQR in R. as noted in Kozumi and Kobayashi (2011), the random walk Metropolis maybe difficult to tune, because a different tuning parameter has to be chosen for every value of τ so as to get an acceptation rate of around 25%.

Kozumi and Kobayashi (2011) propose a location-scale mixture representation of the asymmetric Laplace distribution that allows to find analytical expressions for the conditional posterior densitites of the model. With these tools, they can propose first a conditional natural conjugate prior and second a Gibbs sampler. The merit of the Gibbs sampler is to avoid the specification of a candidate density and of a tuning parameter. The normal – inverted-gamma prior combines nicely with the conditional likelihood in the Gibbs sampler. We can note however that it seems difficult to elicit an informative prior, because we should specify different hyper-

 ISSN 1927-033X

parameters for each quantile. The Gibbs sampler has an important drawback compared to a direct Metropolis approach which is its extreme slowness due to the fact that one has to draw random numbers in an inverted generalized Gaussian for each observation and this is a slow operation which has to be done for each observation separately.

The choice of explanatory variables guided by availability in the LFS data set and includes the traditional variables in Mincerian wage equations plus a few, which in later studies have shown to be significant wage drivers. Our dependent variable in all calculations is the logarithm of wages per hour.

3.3. The model

The formulation we adopt is a standard Mincer equation:

$$\begin{aligned} Ln(y_i) = \alpha_{0\tau} + \beta_{1\tau} Male_i + \beta_{2\tau} exp_i + \beta_{3\tau} exp_i^2 + \beta_{4\tau} union_i + \beta_{5\tau} educ_i + \beta_6 Formal_i \\ + \beta_7 Position_i + \beta_8 Age_i + \beta_9 Industry_i + \epsilon_{i\tau} \end{aligned}$$

Where $(y_i\ ,\ i = 1, \dots, n)$ is the hourly real wage for workers. We have introduced education (number of years), experience and its square, the union status (married/others), formal sector (formal/informal), position (manager/other), age, industry (industry/service) and gender (male/female). Potential experience is calculated as the age minus the assigned years of education minus 6, rounded down to the nearest integer value, min (age – education – 6, age – 18). Education is 1 if worker is less than high school, 2 if the worker reached high school, 3 if worker reached college and 4 if worker reached university. A binary variable for gender, formal sector, position, union status, position, industry (male =1, formal =1, married (union) =1, industry = 1, manager = 1) are also added.

4. Results of empirical analysis

After outlining the methodology, we now provide the results of our empirical analysis. First, we present the data. Then, we discuss the result of Machado and Mata quantile decomposition technique for differences in income densities between workers in formal and informal sectors.

4.1. Empirical data

We employ data from Labor Force Survey (LFS) for Vietnam in 2014. This representative survey is conducted by the General Statistics Office of Vietnam. It is one of the most comprenhensive sources of socio-economic information on Vietnamese households and it plays an important role in the analysis of the living standards of the population.

Our data consist of a sample of 11340 households containing information on household's monthly available income as well as on reference persons' attributes, such as gender, age, education level, experience, status of marital (union), industry (service or manufacturing) and sector (formal or informal), position (manager or basic worker).

Table 1. Hourly wage dispersion for Vietnam 2014

	Min	1st Quarter	Median	Mean	3rd quarter	Max
2014	6.843	9.811	10.062	10.111	10.350	13.346

Table 2. Descriptive statistics for the sample

	Min	1st quarter	Median	Mean	3rd quarter	Max
(Intercept)	8.813	9.030	9.184	9.162	9.274	9.504
Male	0.07279	0.09432	0.09661	0.09765	0.10510	0.11068
Experience	-0.0031	-0.002	0.0024	0.0024	0.0043	0.0091
Experience square	-0.005	-0.003	-0.003	-0.003	-0.003	-0.002
Education	0.1044	0.1153	0.1167	0.1245	0.1323	0.1671
Industry	0.0510	0.0758	0.0993	0.1110	0.1391	0.2006
Union	0.02008	0.04461	0.0466	0.046	0.050	0.061
Position	0.1903	0.2019	0.2061	0.2040	0.2070	0.2108
Formal	-0.1778	-0.1326	-0.1087	-0.1168	-0.0977	-0.0738
Age	0.0160	0.0165	0.0177	0.0183	0.0201	0.0210

	Whole sample	Formal sector	Informal sector
Number of observations	12,058	10,025	2,033
Wage	10.11087 (0.4867)	10.1542 (0.4798)	9.8969 (0.4639)
Male	0.56 (0.4959)	0.5404 (0.4983)	0.6787 (0.4670)
Experience	18.8599 (11.5267)	17.7780 (11.1837)	24.1947 (11.7094)
Experience square	488.5513 (522.2641)	441.1221 (491.6217)	722.4309 (600.2743)
Union	0.6420 (0.4794)	0.6396 (0.4801)	0.6542 (0.4757)
Education	2.0700 (1.1775)	1.4825 (.09102)	17.7780 (11.1837)
Position	0.2010 (0.4007)	0.2258 (0.4181)	0.0787 (0.2693)
Industry	0.4873 (0.4998)	0.5573 (0.4968)	0.4731 (0.4993)
Age	0.4731 (0.4993)	35.6982 (10.4540)	38.2203 (10.7494)

Table 3: Bayesian quantile regression coefficients with a mixture of normal densities on log wages

	Bayes estimate	Lower	Upper	Adj. lower	Adj. upper
(intercept)	-81.844	-105.8248	-68.0850	-131630	131466
Male	0.635	-1.4694	-0.0245	-122382	122383
Experience	8.2046	9.5089	7.1435	25187	25171
Experience square	0.0157	0.0119	0.0217	-830	830
Union	0.2373	0.1265	0.4285	-80022	80021
Education	23.6536	27.5576	20.6107	-141435	141388
Position	7.7186	6.6197	9.4633	-195330	195346
Industry	-0.2574	-0.5944	-0.1050	-146417	146416
Formal	6.4294	5.7029	7.2468	-38168	38181
Age	7.7723	6.8406	8.8693	-9071	9087

The first column summarizes the posterior distribution in one-point estimate, i.e., the Bayes estimate or posterior mean. The second and third column give the 80% credible interval and can be interpreted as “the probability is 80% that the true value of the parameter is in this interval”. The return to experience is much lower than that of education. It is much higher for the first decile and for the median than for the last decile. Being a woman has always meant having a lower wage.

 ISSN 1927-033X

4.2. Decomposition Results:

In our empirical decomposition analysis, the logarithm of the average hourly available income (ln_income) constitutes the outcome variable. Figure 1 illustrates the kernel estimates of the log income densities for the residents of formal and informal sector. The data indicate that the income inequality can be observed. When we look at both distributions, we find that formal workers have the level of the log income higher than the informal workers. The average hourly available income of a person in formal sector was PLN 10.154, whereas for a person in informal sector it was only PLN 9.8

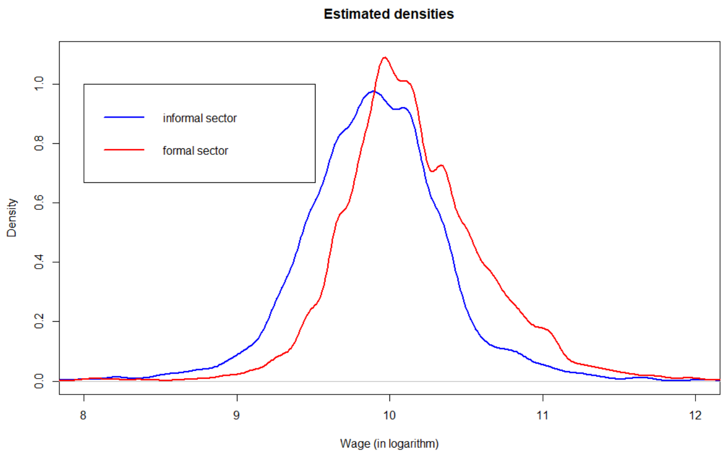

Figure 1: The kernel estimates of the wage densities for the residents of formal and informal sector
(Source: own elaboration based on LFS (2014). Sample averages; standard errors in parentheses)

We establish three explanatory variables in our models: sex – the dichotomous variable encoding gender of a person (number 1 coded the male sex), age – age of a person in years, education – an ordinal variable describing the educational level.

It is useful to look at summary statistics for some covariates (in table 2) among the formal sector there were less men (54%) than it was in the case of the informal sector (67% of men).

Table 4: Estimated decomposition of hourly wage using Machado-Mata decomposition method

Decomposed hourly wage gap					
Hourly wage gap	**Quantile** 10%	**Quantile** 25%	**Quantile** 50%	**Quantile** 75%	**Quantile** 90%
Raw wage gap	0,28	0,22	0,21	0,26	0,34
Explained wage gap	-0,01	-0,01	0,05	0,14	0,30
% explained wage gap (%)	-4,44	-2,37	22,19	54,96	86,56
Unexplained wage gap	0,29	0,23	0,16	0,12	0,05
% unexplained wage gap. (%)	104,44	102,37	77,81	45,04	13,44

(Source: Authors' calculations from a survey of LFS for 2014)

Table 4 also shows that the hourly wage gaps are wider at the top of distribution. Both covariates and coefficients, contribute to the explanation of the total gap sum and their effects are significantly different from zero in all of the estimated deciles (the confidence intervals are not provided because of lack of spae, but they do not include zero). We can see that the impact of coefficients is more important than that of covariates at the bottom of the income distribution. However, the unexplained differential shrinks as we move toward the top of the income distribution. By contrast, the percentage of the explained (die to the characteristics) income differential is considerably greater as we move to the right-side of the distribution.

4.3. Discussions

Firstly, in 2014 in Ho Chi Minh City, the average hourly wage of workers in the informal sector is only 74.28% of the hourly wage of workers in the formal economic sector.

Secondly, the analytical results also show that the main reason for the difference in average wages between workers in the two economic sectors is due to 58.04% explained by the difference in the characteristics of workers and 41.96% explained by the regression coefficient.

Thirdly, when using the percentile regression method, the analysis results show that the wage gap between formal and informal sector workers is lowest in the median and this gap is highest at lowest wage percentile and highest wage percentile.

Fourthly, the result of wage gap separation shows that the wage gap explaination is negative at the wage percentile of 0.10 and 0.25, which provides statistical evidence that: The difference of characteristics of workers in the formal and informal economic sectors do not explain the wage gap between them at the percentile of 0.10 and 0.25. Therefore, for the group of workers with low wages (in the quantile 0.10 and 0.25), the wage difference between workers in the formal and informal sector is merely one reason which is the discrimination in pay of workers.

Finally, the results of analyzation also show that the general trend of wage differences between the two labor groups is: the difference of salary explained (due to the characteristics of the workers) the higher in percentile, the higher in salary; and the difference of salary did not explain (due to regression coefficients) the lower in percentile, the higher in salary.

5. Conclusion

The differences in the wage distribution between formal sector and informal sector have been decomposed into an explained an unexplained component with respect to human capital characteristics. The result of this research show that workers in the informal economy are always more disadvantaged than workers in the formal economic sector. While at the highest wage percentile, the wage gap between employees in the two economic sectors is mainly due to the differences in employee characteristics (explain 85.56%) and at the cash low percentile, the only cause of wage differences is due to the level of remuneration in wage payments (differences due to regression coefficients).

The Machado and Mata decomposition is a very interesting methodology but has the inconvenience of requiring the estimation of a large number of quantile regressions, which is computationally involved. Further research should try to improve on it. A less computer-intensive procedure would be necessary in order to reduce the computation time, particularly for big data sets. Finally, a formal proof of the consistency of this estimator and its asymptotic distribution should also be the object of further research. Machado and Mata (2004) estimate the asymptotic distribution by a simple bootstrap of the generated samples but give no proof that the bootstrap is consistent.

References

Becker, G.S. 1971. The Economics of Discrimination. *Chicago: University of Chicago Press*.

Cain, G. 1986. The economic analysis of labour market discrimination: a survey. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, Amsterdam: North-Holland, 1, 709-730.

Cho, J. and Cho, D. 2009. The Wage Gap in Korea: Analysis of Wage Differences between the Formal and the Informal Sector. *Journal of Asian Economics*, 23:5, 82-97.

Ho Chi Minh City Statistical Office. 2013. 2014 Labor Force Survey Plan, *Ho Chi Minh City*

Dasgupta, S., Bhula-or, R., and Fakthong, T. 2015. Earnings Differentials between Formal and Informal Employment in Thailand. *ILO Asia – Pacific Working Paper Series*.

Koenker, R. and Bassett, J.R. 1978. Regression Quantiles. Economic, 46:1, 33-50.

Lam. L. B., Ha, N. M. and Huong, L. V. 2015. Oaxaca-Blinder model in economic analysis. *Science Magazine HCMC Open University*, 2:41, 3-11.

Machado, J., and Mata, J. 2005. Countterfactual decomposition of changes in wage distributions using quantile regession. *Journal of applied Econometrics*, 20, 445-465.

Mincer, J. 1974. Schooingl, Experience and Earning, Nation bureau of economic research. *Comlombia University Press.*

Nguyen Huu Chi. 2010. Employment for rural-to-urban migrant workers: analyzes of employment options and income of migrant workers compared to in-urban labor in the Red River Delta. *Hanoi: Tri Thuc Publishing House.*

Nguyen Thanh Binh. 2014. Differences between labor in formal and informal economic sectors in Ho Chi Minh City, Master's thesis*, Ho Chi Minh City Open University*.

Roxana Maurizio 2010. Lao động phi chính thức và nghèo đối ở châu Mỹ Latin. Trường hợp của Argentina, Brazil, Chile và Peru. *Hà Nội: Nhà xuất bản Tri Thức.*

Roxana Maurizio. 2010. Level of satisfaction with informal jobs and employment in the informal sector in Vietnam: economic implications and public policy. *International Conference, Hanoi: Knowledge Publishing House.*

Tansel, A. and Kan, E.O. 2012. The Formal/Informal Employment Earnings Gap: Evidence from Turkey. *IZA Discussion Paper* 6271

General Statistics Office. 2013. Guidelines for LFS 2014. Hanoi: Statistical Publishing House.

General Statistics Office. 2016. Results of the Labor force survey in 2011-2015. *Hanoi: Statistical Publishing House.*

Tran Thi Tuan Anh. 2015. Applied regression analysis of wage gap analysis in Vietnam, PhD thesis, *Ho Chi Minh City University of Economics.*

Zuo, H. 2013. The Earnings Differential between Formal and Informanl Employees in Urban China. *Contemporary Economics*, 7, 115- 124.

 ISSN 1927-033X