I Z A

Who Suffers the Penalty?
A Panel Data Analysis of Earnings Gaps in Vietnam

Huu Chi Nguyen
Christophe J. Nordman
François Roubaud

Who Suffers the Penalty? A Panel Data Analysis of Earnings Gaps in Vietnam

Huu Chi Nguyen
Université Paris Nord, DIAL

Christophe J. Nordman
IRD, DIAL and IZA

François Roubaud
IRD, DIAL

ABSTRACT

Who Suffers the Penalty?
A Panel Data Analysis of Earnings Gaps in Vietnam[*]

In spite of its predominant economic weight in developing countries, little is known about the informal sector earnings structure compared to that of the formal sector. Taking advantage of the VHLSS dataset in Vietnam, in particular its three wave panel data (2002, 2004, 2006), we assess the magnitude of various formal-informal earnings gaps while addressing heterogeneity at three different levels: the worker, the job (wage employment vs. self-employment) and the earnings distribution. We estimate fixed effects and quantile regressions to control for unobserved individual characteristics. Our results suggest that the informal sector earnings gap highly depends on the workers' job status and on their relative position in the earnings distribution. Penalties may in some cases turn into premiums. By comparing our results with studies in other developing countries, we draw conclusions highlighting the Vietnam's labour market specificity.

JEL Classification: J21, J23, J24, J31, O17

Keywords: informal employment, earnings gap, transition matrix, panel data, Vietnam

Corresponding author:

Christophe J. Nordman
DIAL
4 rue d'Enghien
75010 Paris
France
E-mail: nordman@dial.prd.fr

---

1. Introduction

In spite of its predominant economic weight in developing countries, little is known about the informal sector's earnings structure compared to that of the formal sector. Some works have been carried out in this field using household surveys, but they only consider some emerging Latin American countries (Argentina, Brazil, Colombia and Mexico; Gong *et al*., 2004; Perry *et al*., 2007) and, more recently, South Africa, Ghana and Tanzania for Africa (Falco *et al*., 2010; Bargain and Kwenda, 2011). As a matter of consequence, there is still no way to generalise these (diverging) results to other parts of the developing world, in particular in countries where the informal sector is the most widespread, that is Sub-Saharan African and more generally very poor countries.

From a labour market perspective, two competing views regarding informality are at stake in the literature: the exclusion and the exit hypotheses, following Hirschman's seminal work (Perry *et al.*, 2007). The first one, also called the 'dualist approach', is an extension of the works by Lewis (1954) and Harris and Todaro (1970). It is based on a dual labour market model where the informal sector is considered as a residual component and is totally unrelated to the formal economy. It is a subsistence economy that only exists because the formal economy is incapable of providing enough jobs, and is condemned to disappear with the development process. Informal workers, suffering from poor labour conditions, are queuing for better jobs in the formal sector. The second view, also known as the 'legalist approach' considers that the informal sector is made up of micro-entrepreneurs who prefer to operate informally to evade economic regulations (de Soto, 1989); this conservative school of thought is in sharp contrast to the former in that the choice of informality is voluntary due to the exorbitant legalisation costs associated with formal status and registration.

Empirical evidence shows, however, that the real situation is a mix of these two hypotheses. Confirming Fields (1990)'s stylized assessment, some studies stress the huge heterogeneity among informal jobs, which combine two main components (Roubaud, 1994; Maloney, 1999, 2004; Perry *et al.*, 2007; Bargain and Kwenda, 2011): a lower-tier segment, where occupying an informal job is a constraint choice ('exclusion hypothesis'); an upper-tier segment, in which informal jobs are chosen for better earnings, and non-pecuniary benefits ('exit hypothesis'). Usually, the former segment is assimilated to the informal wage jobs,

while the latter is associated with the self-employed jobs. Therefore, whether one segment is predominant over the other is an empirical question, depending on local circumstances. To test these alternative views, one major strand of literature focuses on the estimation of earnings gaps. Embedded in reveled preferences principle, and considering income as a proxy of individual utility, this approach assumes that if informal workers earn more than their formal counterparts, all else being equal, this would reflect a deliberate choice of the former to be informal workers. This may not be true for all informal workers. Thus, the challenge is to identify segments of jobs or positions in the income distribution where informal workers get a higher pay. This is the method we follow in this paper in the case of Vietnam.

Taking advantage of a three-wave panel dataset (VHLSS 2002-2004-2006), we ask the following questions: Is there an informal job earnings penalty? Do some informal jobs provide pecuniary premiums? Do possible gaps vary along the earnings distribution? The case of Vietnam is interesting because it has experienced spectacular social, economic and political changes in the recent period. The growth model embraced by Vietnam during the last two decades has prompted deep social economic transformation. The private sector has been thriving with the transition of a centrally planned economy towards a 'socialist-oriented market economy' since the *Doi Moi* (Renovation) launched in 1986 (see Migheli, 2012). Economic growth helped reduce poverty considerably but, in the meantime, increased social inequality. The gap within a region and between urban and social areas widened (VASS, 2010; Cling *et al.*, 2009). On the labour market, two main striking features are at stake in recent years.

First, Vietnam's impressive economic growth has triggered a sharp increase in the rate of wage employment: from 19% in 1998 to 33% in 2006 (Table 1). This spreading of wage employment, accompanied by a steep decline in agricultural jobs, has affected all population categories (urban, rural, male, female, skilled and unskilled), but substantial differences in level subsist. Wage employment is obviously more developed among the most skilled manpower (86% among the highly skilled as opposed to barely one-quarter among the unskilled; figures not shown), and it is also more prevalent among urban dwellers and among men (35% compared to 25% for women). This trend is due to a vibrant urbanization process. But, at the same time, in all kinds of geographic areas, the proportion of out farm jobs has been on the rise, a shift particularly important in suburban areas (Cling *et al.*, 2010b).

Table 1. Changes in labour structure and earnings in Vietnam, 1998-2006

Sector	Jobs (%)				Real earnings* (100 = 1998; wage only)			
	1998	2002	2004	2006	1998	2002	2004	2006
Agriculture	67.1	56.5	52.0	49.2	100	96.2	107.4	128.3
Secondary sector	13.9	19.7	21.7	23.0	100	109.4	119.6	134.3
Services	19.0	23.8	26.3	27.8	100	146.1	158.3	177.7
Total	100	100	100	100	100	121.2	137.1	155.7
Wage workers	17.5	28.6	31.0	33.1	-	-	-	-

Source: VLSS1998, VHLSS 2002, 2004& 2006, GSO; authors' calculation.

Note: Secondary sector includes fishery, mining, manufacture and construction.

*: regional deflators and provincial CPIs have been elaborated to compute real earnings (see details in text and Box 2 in Appendix).

The second important feature of the Vietnamese labour market is that wages gradually rose from 1998 to 2006 (Table 1). Sharp economic growth prompted a 56% increase in wage earners' average annual remuneration over the period observed, which works out at an average annual growth rate of 5.7%. Wage dynamics was higher for the semi-skilled and high skilled workers than for unskilled workers (67%, 62% and 36% respectively; figures not shown).

This on-going restructuring of the labour market clearly benefitted the non-farm private sector. Although the formal sector of the economy has grown progressively, the steady dynamics and still important share of informal employment along with the increasing share of wage workers in total employment represent intriguing facets of the Vietnamese labour market. The Labour Force Survey conducted in 2007 gives a precise picture of the informal economy (Cling *et al.*, 2010a): Informal *sector* jobs represent 23% of total jobs and nearly a half of non-farm jobs; informal jobs, those defined at the *worker* level, account for 82% of total jobs and two-thirds of non-farm jobs.

While most of the papers analysing informal sector earnings gaps are drawn from Latin American or African countries, Vietnam therefore represents an interesting case. Do Vietnam's specific circumstances (unique economic regime, role of the State, fast growing economy) make a difference? This topic is all the more relevant that it is directly linked to a key policy issue: a universal social insurance scheme is to be implemented in the coming years by the Socio-Economic Development Strategy (SEDS 2011-2020). Does the affiliation of workers to social insurance really matter to them in obtaining higher earnings? Our objective is thus to shed light on the above-mentioned alternative views about the informal economy using the formal/informal earning gaps approach. We also intend to complete the

puzzle by broadening the spectrum of developing countries where the alternative views can be assessed, in order to draw more general conclusions. In particular, does the exit option still hold in poorer countries? Of course, by this means, we do not provide a formal test of the theoretical hypotheses, but we intend to enlighten the possible mechanisms at work using earnings gaps as reasonable indicators of differences in job quality.

Our analysis consists of estimating OLS and quantile earnings regressions. In particular, we estimate fixed effects quantile regressions to control for unobserved individual characteristics, focusing particularly on heterogeneity within both the formal and informal employment categories. We address the important issue of heterogeneity at three levels: the worker, taking into account individual unobserved characteristics; the job, comparing wage workers with self-employed workers; the earnings distribution. Thus we expand in various ways Rand and Torm (2012a), the closest paper to our approach.[1] These authors also estimate formal-informal earnings gaps, which are complemented in their study by decomposition techniques. Their work is partially limited by the survey characteristics they use. First, their sample is not representative of the informal sector, as their survey only captures its upper-tier segment in some provinces (informal manufacturing firms with professional premises). Second, relying on an enterprise survey, their variable of interest is restricted to the average wage per firm, which does not allow them to investigate individual heterogeneity, in particular distributional effects. Third, they are not able to control for time-invariant individual characteristics due to the cross-sectional nature of their dataset.

Our results suggest that the informal earnings gap highly depends on the workers' job status and on their relative position in the earnings distribution. Penalties may in some cases turn into premiums. In particular, while informal workers suffer penalties vis-à-vis formal workers, this feature is mainly due to informal wage earners. In fact, in comparison with formal wage workers, informal self-employed workers receive a premium which is increasing along the pay ladder.

The remainder of this paper is as follows. Section 2 presents the data and descriptive statistics. Section 3 focuses on the econometric approach. Results on earnings gaps are discussed in Section 4. Section 5 concludes.

---

[1] The few existing papers on Vietnam addressing informality used various approaches: job satisfaction, determinants and consequences of business registration, wage gaps (Cling *et al*., 2012; Razafindrakoto *et al*., 2012; Rand and Torm, 2012a, 2012b).

2. Data and Descriptive Statistics

The data used in this paper are drawn from three successive rounds of the Vietnam Household Living Standards Surveys (VHLSS 2002, 2004 and 2006). These surveys are LSMS surveys' type, probably one of the most popular household surveys in developing countries. In addition, the VHLSS has the reputation of being one of the best LSMS surveys in the world.

Construction of the panel

The sample size of the VHLSSs is quite large even if it has been progressively reduced, from 75,000 in 2002 to 45,000 in 2004 and 2006. A detailed questionnaire (including expenditures and other subject specific modules) has been applied to a random subsample of 30,000 individuals and around 9,000 households respectively. To track individual changes over time, a panel component has been implemented, selected among the three subsamples. Individuals have been matched between the three surveys using the common individual identifier across years, cross-checked with gender, age and other individual information (see the panel construction in Box 1 in Appendix). Retaining only those individuals who are 15 years old or more and engaged in non-agricultural and non-public activities, our analysis is based on a panel of non-farm workers including 948 individuals observed in all three years (balanced panel). In the unbalanced parts, there remain 1,550 individuals observed in both 2002 and 2004 but not in 2006, and 1,897 individuals who are observed as non-farm workers in both 2004 and 2006, but were not surveyed in 2002.

As in any panel data analysis, potential selective attrition should be considered and addressed. The VHLSS sample design consists in a rotating panel, half of the households being randomly renewed from one round to the other. Comparisons of means and distributions of earnings and observables between the cross-sectional samples and the panel sub-sample suggest that selective attrition is not an issue. This finding is consistent with previous papers using the panel component of the VHLSSs, even if, to our knowledge, we are the first to rely on a three point panel analysis (2002-2004-2006), while other studies are based on two consecutive rounds of data: 2002-2004 and/or 2004-2006 (Vijverberg *et al.*, 2006; Tran Quoc Trung and Nguyen Thanh Tung, 2008; Oostendorp *et al.*, 2009; Imbert, 2011).

The VHLSS does not allow us to capture the concept of informal sector following the international definition strictly (ILO, 2003)[2], as the survey has not been designed for such a purpose. In Vietnam, the *informal sector* is defined as all private unincorporated enterprises (that is not considered under the Law on Enterprises, 2000, 2005) that produce at least some of their goods and services for sale or barter, are not registered (have no business licence) and are engaged in non-agricultural activities. The *informal employment* corresponds to employment with no social security insurance (Cling *et al.*, 2010a). On the job side in the VHLSS, the formal/informal divide can only be computed for wage workers. On the firm side, household businesses can be split between registered and not registered ones, but no information is available on the jobs generated by these businesses. Therefore, we created an informality proxy, still consistent with the ILO definition of informal employment, which combines job and firm approaches. Four main groups are distinguished. Among wage workers, informal ones are those who do not benefit from social security.[3] Among employers and self-employed, informal workers are those whose business is not registered. As in other studies, we exclude agriculture from the analysis. This classification provides the best available measures of informality in Vietnam, previous to the LFS 2007 (which unfortunately does not provide any panel component).

Apart from our informality variable, we compute the labour income associated with each remunerated job. For wage workers, earnings are obtained by summing the direct wage with all the supplementary benefits perceived in cash or in kind and converted into pecuniary equivalent (public holidays, bonuses, social allowance, etc.). For the self-employed, we compute their annual net income by subtracting all the expenses engaged (intermediary consumption, labour costs, taxes) to the production generated by the household business. Hourly earnings used in the econometric analysis are deduced using the total number of hours worked per year. Regional and time deflators have been elaborated to compute real earnings (see Box 2 in Appendix). Finally, to reduce a possible bias due to measurement and reporting errors in the earnings and explanatory variables, we trim the data and drop influential outliers and observations with high leverage points that we identify by the

---

[2] According to the resolution of the 15th International Conference of Labour Statisticians, the *informal sector* is broadly defined as consisting of all unregistered or micro unincorporated enterprises. The ILO more recently developed a broader concept of *informal employment* that takes into account the characteristics of the jobs: social security coverage, existence of a written contract or pay slip, paid leave, etc.

[3] As the VHLSSs do not provide detailed information on labour contract as well as on social security affiliation in the employment module, we base our definition using a question asking whether each household member has benefited from health insurance, which is available in a specific module on health care.

DFITS-statistic. As suggested by Belsley, Kuh and Welsch (1980), we use a cutoff-value $|DFITS|_{ihj} > 2\sqrt{k/N}$ with k, the degrees of freedom (plus 1) and N the number of observations. This procedure removes 497 observations from our initial unbalanced panel sample of 9,571 observations.[4] The number of removed observations is the largest in 2004 (respectively 233; 137 and 127 for 2004, 2006 and 2002).

Descriptive statistics and validity checks

Table 2 presents some basic summary statistics of the main characteristics of our panel. These statistics are reported for the sub-samples of wage/self-employed workers, broken down by formal and informal jobs.

The results obtained for average earnings are in line with common findings. Workers holding formal jobs earn more on average than those engaged in informal jobs. Among each group of formal and informal workers, self-employed workers are those with higher earnings in comparison with wage earners. Informal workers tend to be younger than their formal worker counterparts, especially for wage workers. Self-employed workers exhibit on average longer potential experience in the labour market (which is calculated as age minus years of reported schooling minus five). As expected, workers having higher level of education are less likely to be engaged in informal employment and vice versa.

At the aggregate level, the gender ratio does not vary between formal and informal jobs. However, female workers have more opportunity to get formal wage jobs than informal ones. Finally, formal and informal workers are differently allocated across branches of activity. Specifically, informal employment is found more in trade, restaurants and transportation, while formal jobs are more concentrated in services. Interestingly, the share of manufacture is much higher for informal jobs than for formal ones (31% vs. 18%). Within employment sectors, the distribution is also fairly unbalanced: formal wage workers are stubbornly engaged in services (60%), whereas formal self-employed workers hold transportation and hotel & restaurant jobs (12% and 52% respectively). Informal wage workers engaged prominently in construction (13%) and trade (35%) while informal self-employed job's structure looks like the formal self-employed one. These significant differences in the distribution of job structure underline the importance of controlling for sectors of activity in our earnings estimations.

---

[4] The regression analysis drops 167 observations out from the initial unbalanced panel sample of 9,738 observations due to missing values in the set of regressed variables.

Table 2. Summary Statistics (pooled waves 2002-2004-2006)

	Formal workers						Informal workers					
	All workers		Self-Employed		Wage workers		All workers		Self-employed		Wage workers	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Hourly earnings	1.949	0.63	2.134	0.70	1.885	0.60	1.552	0.60	1.645	0.67	1.450	0.50
Potential experience	21.98	10.65	26.12	10.66	20.52	10.26	24.22	13.08	28.36	13.08	19.66	11.48
Age	38.79	10.13	40.47	9.93	38.20	10.14	37.05	12.22	40.88	11.98	32.83	11.03
Female	0.462	0.50	0.536	0.50	0.436	0.50	0.438	0.50	0.603	0.49	0.256	0.44
Married	0.787	0.41	0.824	0.38	0.775	0.42	0.703	0.46	0.793	0.41	0.604	0.49
Position in the family												
Head of household	0.399	0.49	0.390	0.49	0.402	0.49	0.376	0.48	0.382	0.49	0.369	0.48
Spouse	0.299	0.46	0.375	0.48	0.273	0.45	0.257	0.44	0.386	0.49	0.115	0.32
Children	0.273	0.45	0.199	0.40	0.299	0.46	0.337	0.47	0.209	0.41	0.479	0.50
Others	0.029	0.17	0.036	0.19	0.026	0.16	0.030	0.17	0.023	0.15	0.036	0.19
Years of education	11.80	3.52	9.31	3.18	12.68	3.20	7.83	3.32	7.53	3.34	8.17	3.29
Industry												
Food and beverage	0.037	0.19	0.038	0.19	0.037	0.19	0.060	0.24	0.068	0.25	0.051	0.22
Textile, leather, wood, hand.	0.075	0.26	0.068	0.25	0.077	0.27	0.154	0.36	0.147	0.35	0.162	0.37
Construction	0.119	0.32	0.062	0.24	0.139	0.35	0.083	0.28	0.040	0.20	0.131	0.34
Whole sale	0.031	0.17	0.004	0.06	0.041	0.20	0.180	0.38	0.013	0.11	0.364	0.48
Retail sale	0.025	0.16	0.049	0.22	0.016	0.13	0.017	0.13	0.025	0.16	0.009	0.09
Hotel and restaurant	0.149	0.36	0.513	0.50	0.022	0.15	0.264	0.44	0.433	0.50	0.078	0.27
Transportation & wharehouse	0.040	0.20	0.134	0.34	0.006	0.08	0.086	0.28	0.135	0.34	0.033	0.18
Other manufacture	0.059	0.24	0.067	0.25	0.057	0.23	0.093	0.29	0.093	0.29	0.092	0.29
Other services	0.465	0.50	0.066	0.25	0.605	0.49	0.063	0.24	0.047	0.21	0.081	0.27
Number of observations	4,036		1,049		2,987		5038		2,639		2,399	

Source: VHLSS, 2002, 2004 & 2006, GSO; authors' calculations.

Table 3 reports the transition matrices of employment status between 2002-2004, 2004-2006 and 2002-2006 obtained from the unbalanced panel dataset of all individuals aged 15 or more. The categories shown in the matrices include then not only the four non-farm employment statuses but also "agriculture" and "not-working" (the latter category including, simplifying the notation, those who are inactive or unemployed). This presentation allows identification of both transition flows within the non-farm sector employment and those into or out of the non-farm sector. The figures in the first two rows and columns of each matrix reveal that the latter are not negligible. Among these, we observe that the most important flows are those between informal non-farm and agricultural jobs. These patterns of mobility would partly reflect the low entry barriers to both sectors as well as the fact that the majority of the workforce in Vietnam is still predominantly employed in agriculture. Another striking

finding is the rather high probability of becoming inactive or unemployed for those who were previously self-employed.

Table 3. Transition matrices of employment status (%)

2002	2004						Total
	Not-working	Agri-cultural emp.	Formal Wage	Informal Wage	Formal Self-employed	Informal Self-employed	
Not-working	62.77	19.64	3.45	5.74	2.33	6.06	100 (21.7)
Agricultural emp.	7.4	80.39	1.59	5.58	0.68	4.36	100 (50.2)
Formal Wage worker	3.51	6.15	74.34	13.76	0.88	1.37	100 (8.9)
Informal Wage worker	5.01	17.05	3.65	62.25	3.11	8.93	100 (6.5)
Formal Self-employed worker	7.10	5.03	1.78	5.33	55.92	24.85	100 (2.9)
Informal Self-employed worker	6.70	16.3	1.45	8.06	12.23	55.25	100 (9.8)
Total	18.88	47.98	8.64	10.25	3.96	10.28	100 (100)
2004	2006						
Not-working	81.34	10.36	1.96	3.5	0.92	1.92	100 (33.4)
Agricultural emp.	10.33	78.24	1.54	4.73	0.89	4.26	100 (39.7)
Formal Wage worker	3.72	5.37	81.51	6.61	0.62	2.17	100 (6.7)
Informal Wage worker	5.5	15.26	8.7	60.16	2.57	7.81	100 (7.9)
Formal Self-employed worker	12.13	6.07	1.57	3.72	49.12	27.4	100 (3.6)
Informal Self-employed worker	7.70	14.68	1.84	6.58	9.3	59.9	100 (8.7)
Total	33.07	37.61	7.66	8.92	3.46	9.27	100 (100)
2002	2006						
Not-working	55.29	17.65	6.87	9.62	2.64	7.93	100 (19.9)
Agricultural emp.	7.32	76.44	2.76	6.44	1.12	5.92	100 (52.5)
Formal Wage worker	4.29	5.81	74.49	11.87	0.76	2.78	100 (8.3)
Informal Wage worker	5.35	17.06	4.35	55.18	3.68	14.38	100 (6.2)
Formal Self-employed worker	13.10	7.59	2.07	2.07	51.03	24.14	100 (3.2)
Informal Self-employed worker	10.34	20.46	2.95	7.38	9.7	49.16	100 (9.9)
Total	16.95	47.48	9.64	10.55	3.93	11.45	100 (100)

Source: VHLSS, 2002 & 2006, GSO; authors' calculations.
Note: the number of observations for the three transition matrices (2002-2004, 2004-2006 and 2002-2006) are 11,425; 14,348 and 4,758 respectively.

For the purpose of measuring earnings gaps, we focus hereafter on the transition flows within the non-farm jobs (shaded cells in the transition matrices). They show that, on average, not negligible flows are observed between non-farm job's categories. For the two time periods, around one quarter of workers changed position from one of our four job's status to another. Around 20% of the total sample moved from informal to formal jobs and

the rates of formal-informal transitions are about 40%. However, the flows are balanced in absolute terms. The fluidity between wage and non-wage jobs is smaller, but is far from negligible (from 13% to 15% of the total sample, depending on the years). Here again, the movements to and from wage jobs are relatively symmetrical. At a more disaggregated level, job mobility is at its highest for formal self-employed workers, where less than two thirds keep the same status in our different panels. Formal wage workers are the most stable (82% to 74% of stayers), while informal workers are in between with a proportion of stayers ranging from 55% to 62%. Formal wage workers mainly move to informal wage jobs. When moving, informal wage workers tend to privilege self-employed and formal wage jobs. Formal self-employed movers mainly get their business informalised (probably due to adverse conditions). A lower share of informal self-employed workers makes the reverse move, by formalising their business. However, a substantial proportion also closes their business to become informal wage workers.

On the methodological side, the substantial numbers of movers is a key for our estimation strategy (see next section). It is also important that the movements between these types of jobs do not specifically concentrate on certain workers in the sample. In order to verify this pattern of labour mobility, we examine the rate of transitions across employment statuses at different earnings quantile levels. Figure 0 in Appendix shows the proportion of movers out of (or into) informal employment in each earnings quantiles in the base (or current) period. As can be seen in this figure, there is a substantial proportion of movements in both directions at all earnings levels. Overall, the transitions are more frequent in the upper quintiles and this is found for both types of transition.

Another necessary validity check for our estimation of earnings gaps is to verify the existence of actual job changes resulting from declared changes in occupation and industry type. Theoretically, tenure in the current job could be another ideal criterion to assess job changes. However, available information in VHLSS 2002 does not allow this assessment. Table A0 in Appendix reports the rates of changes in occupation and/or industry of activity accompanied by transitions across informal and formal employment. Overall, around 60% of inter-sector movements are concomitant with changes in at least one of the two employment characteristics. This reinforces our confidence in the existence of limited measurement errors related to the reported employment status at certain time. All in all, the high consistency between transition matrices over different samples and time periods appears to be a sound indicator of data quality. We would claim that the observed changes reflect real phenomena and do not mainly capture measurement errors.

Table 4. Earnings dynamics by employment status between 2004 and 2006

Employment status in 2004	Real earnings levels in 2006				
	Employment status in 2006				
	Formal wage worker	Informal wage worker	Formal self-employed	Informal self-employed	**Total**
Formal wage worker	100.0	58.5	106.8	83.1	75.0
Informal wage worker	62.5	49.7	72.7	55.2	54.5
Formal self-employed	99.8	67.6	157.7	132.7	114.8
Informal self-employed	86.2	61.4	123.5	94.8	85.3
Total	82.9	59.2	121.7	88.0	80.1
	Real hourly earnings growth 2004-2006				
Employment status in 2004	Employment status in 2006				
Formal wage worker	100.0	90.0	100.2	95.7	94.5
Informal wage worker	92.4	86.1	94.9	87.5	87.9
Formal self-employed	99.7	92.9	104.5	102.4	100.9
Informal self-employed	97.5	90.8	101.6	97.7	96.4
Total	96.9	90.0	102.1	95.5	95.0

Source: VHLSS 2002, 2004 & 2006, GSO; authors' calculations.
Note: base 100=Income level and income growth compared to formal wage workers' stayers between 2004 and 2006. For example, considering the upper panel, compared to formal wage worker stayers, those who were formal wage workers in 2004 and changed to formal self-employment in 2006 had on average an earning level 6.8% higher. The number of observations in the wide form panel used for computing earnings dynamics is 2,827.

To end this section on descriptive analysis, we look at the earnings dynamics by employment status, focusing on the period 2004-2006. The first panel of Table 4 shows the level of real earnings in 2006 by transition status, formal wage stayers being our basis. Consistently with Table 2, formal self-employed workers get the highest pay, while informal wage workers are at the lowest end of the earnings ladder. Compared to the pooled sample, in 2006, informal self-employed workers reversed their position with formal wage workers, meaning that the earnings hierarchy between these two categories of workers is not fixed, but may vary over time. Furthermore, earnings levels are highly dependent on transitions. For instance, and as expected, whatever their job status in 2004, those who moved to informal wage jobs earn the less. Conversely, the workers who got the opportunity to open a formal business earn the most. The results are quite similar in terms of earnings growth (second panel of Table 4). Systematically, moving to informal wage jobs is associated with the lowest increase in earnings over the period, whereas being able to change to a formal self-employed job is associated with the highest earnings growth. Of course, these unconditional averages should be controlled for observed and unobserved characteristics, which is the purpose of the following sections. Furthermore, changes in job states are not systematically associated with upwards (or downwards) trends in incomes. Among our twelve groups of movers, six of them suffered a lower income growth than the respective stayers, while four benefited from a relative increase (the two remaining groups did not

register any significant changes). This reinforces the following identification strategy of earnings gaps based on movers and stayers.

3.   Econometric Approach to Measuring Informal-Formal Earnings Gaps

The analysis consists of assessing the magnitude of different informal-formal earnings gaps using OLS and quantile regressions with log hourly earnings as dependent variable.[5] Standard earnings equations are thus estimated at the mean and at various conditional quantiles of the earnings distribution. The models are regressed on a pooled sample of workers over years employed formally and informally. The different covariates are the completed years of education, the years of potential experience (with quadratic profiles for these two regressors), a dummy for being married, a dummy for being a woman, eight dummy variables of branch activities[6], seven regional dummies and two time dummies to control for macroeconomic trend effects on earnings.

A number of studies based on data on African manufacturing firms have shown that wages are positively correlated to firm size, conditional on standard human capital variables.[7] The literature discusses numerous reasons why wages are positively correlated with firm size. One of the frequently made arguments is that firm size is correlated with omitted worker quality because large firms usually attract more productive workers. In this paper, due to lack of information on the demand side characteristics, we cannot control for the size of the wage workers' firms but we control for both observed human capital and time-invariant unobserved characteristics, thus mitigating the drawback of not accounting for firm size in the regressions.

To account for informal-formal differences in earnings at the mean earnings level, we rely on Fixed Effects regressions (FE) accounting for time-invariant unobserved heterogeneity.

---

[5] One could object that *hours* of work are likely to be measured with serious error which would contaminate the estimated coefficients and the inferences about who earns more than who else. An alternative would then be to use log *monthly* earnings in place of log *hourly* earnings. We investigated this path and obtained qualitatively very similar findings, in particular in the fixed effects estimates. Yet, we believe that hourly earnings have the crucial advantage of being able to account for differences in labour effort across individuals, in particular across activity sectors, branches, or gender.

[6] These dummies include "Food and beverage", "Textile, leather, wood, handicraft", "Construction", "Whole sale", "Retail sale", "Hotel and restaurant", "Transportation and wharehouse" and "Other manufacture" (the reference being "Public sector services").

[7] See Strobl and Thornton (2004) and Söderbom *et al.* (2005).

The FE model can be written as

$$y_{it} = x_{it}'\beta + \gamma I_{it} + \alpha_i + \varepsilon_{it} \qquad (1)$$

where x_{it} denotes the vector of characteristics of individual i observed at time t (which includes a constant term), I_{it} represents a dummy taking value one if person i observed at time t is an informal worker. α_i is the individual fixed effect and ε_{it} is an i.i.d. normally distributed stochastic term absorbing measurement error. Note[8] that $E[\varepsilon_{it}|x_{it}, I_{it}, \alpha_i] = 0$.

The estimated coefficient $\hat{\gamma}$ is interpreted as a measure of the conditional earnings premium/penalty experienced by workers who change status between informal to formal jobs (or the reverse). However, as informal employment is highly heterogeneous, a finer job divide should be considered. We define four categories of workers split by job status (wage workers vs. self-employed workers) and institutional sector (formal vs. informal) and create four dummies taking value one if the individual i at time t is an informal wage worker (IW_{it}), a formal wage worker (FW_{it}), an informal self-employed worker (IS_{it}) and a formal self-employed worker (FS_{it}). Taking the formal wage workers as the reference category, the model we estimate is written as:

$$y_{it} = x_{it}'\beta + \delta IW_{it} + \theta IS_{it} + \lambda FS_{it} + \alpha_i + \varepsilon_{it} \qquad (2)$$

The estimated coefficients $\hat{\delta}$, $\hat{\theta}$ and $\hat{\lambda}$ are interpreted, respectively, as the *IW – FW*, *IS – FW* and *FS – FW* conditional earnings gaps. Identification of these conditional earnings gaps relies on the presence in the sample of *movers* between employment states over time. Those movers can be compared to the *stayers* in terms of earnings. As an illustration, we consider a simple two-period example and eight cases of transitions out of the various possibilities of professional trajectories (which are 16 in a two-period example):

2 cases of stayers:

$$E[y_{i2} - y_{i1}|IW_{i1} = 1, IW_{i2} = 1] = \Delta \qquad (3)$$
$$E[y_{i2} - y_{i1}|IS_{i1} = 1, IS_{i2} = 1] = \Delta \qquad (4)$$

---

[8] One could use a random effect (RE) model assuming in addition that $E[\alpha_i/x_{it}, I_{it}] = 0$. However, as in many other cases, this condition is very unlikely to be satisfied as individual unobserved characteristics are generally correlated with workers' observable characteristics. Hausman's specification test indeed confirmed a systematic difference in the FE and RE estimators.

With $\Delta = (x'_{i2} - x'_{i1})\beta$

6 cases of movers:

$$E[y_{i2} - y_{i1}|IW_{i1} = 1, IS_{i2} = 1] = \Delta + \theta - \delta \qquad (5)$$

$$E[y_{i2} - y_{i1}|IW_{i1} = 1, FW_{i2} = 1] = \Delta - \delta \qquad (6)$$

$$E[y_{i2} - y_{i1}|FW_{i1} = 1, IS_{i2} = 1] = \Delta + \theta \qquad (7)$$

$$E[y_{i2} - y_{i1}|FW_{i1} = 1, FS_{i2} = 1] = \Delta + \lambda \qquad (8)$$

$$E[y_{i2} - y_{i1}|IS_{i1} = 1, FS_{i2} = 1] = \Delta + \lambda - \theta \qquad (9)$$

$$E[y_{i2} - y_{i1}|IS_{i1} = 1, FW_{i2} = 1] = \Delta - \theta \qquad (10)$$

With $\Delta = (x'_{i2} - x'_{i1})\beta$

Equations (3) and (4) give examples of the changes in earnings for *stayers*, that is for workers that do not change their employment state between the two periods. Equations (5) and (6) illustrate the changes in earnings for those workers coming from an informal wage job and moving, respectively, into an informal self-employed job and a formal wage job; equations (7) and (8) represent these earnings differentials for those coming from a formal wage employment and moving, respectively, into an informal self-employed job and a formal self-employed job. Finally, the cases of informal self-employed workers moving to, respectively, formal self-employed and formal wage jobs are considered in equations (9) and (10).

This identification strategy of FE on movers is quite standard but, in practice, one should verify that the number of moves across employment states is sufficient for a valid use of this estimator. We verified that this was the case in Table 3. More generally, the identification strategy supposes that movers change more or less randomly between employment states, or at least that they do not systematically move for better earnings. However, people may change jobs in particular if they see an opportunity to earn more. The earnings matrix (Table 4) shows that this is actually not the case.

Finally, to allow the earnings gaps between job statuses to differ along the earnings distribution, we rely on Quantile Regressions (QR). Quantile earnings regressions 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. Using our previous notation, the model that we seek to estimate is:

$$q_\varrho(y_{it}) = x'_{it}\beta(\varrho) + \delta(\varrho)IW_{it} + \theta(\varrho)IS_{it} + \lambda(\varrho)FS_{it} + \alpha_i, \forall \varrho \in [0,1] \tag{11}$$

Where $q_\varrho(y_{it})$ is the ϱth conditional quantile of the log hourly earnings. The set of coefficients $\beta(\varrho)$ provide the estimated rates of return to the different covariates at the ϱth quantile of the log earnings distribution and the coefficients $\delta(\varrho), \theta(\varrho)$ and $\lambda(\varrho)$ measure the parts of the earnings differentials that are due to informal-formal job differences at the various quantiles.

We then turn to Fixed Effects Quantile Regressions (FEQR). The extension of the standard QR model to longitudinal data has been originally developed by Koenker (2004). Recently, Canay (2011) proposed an alternative and simpler approach which assumes that the unobserved heterogeneity terms have a pure location shift effect on the conditional quantiles of the dependent variable. In other words, they are assumed to affect all quantiles in the same way. It follows that these unobserved terms can be estimated in a first step by traditional mean estimations (for instance by FE OLS). Then, the predicted $\hat{\alpha}_i$ are used to correct earnings, such as $\hat{y}_i = y_i - \hat{\alpha}_i$, which are regressed on the other regressors by traditional QR.

Finally, when running the regressions (2) and (11), we provide robust standard errors using bootstrap replications.

4. Earnings Gaps Analysis

In this section, we discuss the earning gaps between formal and informal jobs. At the aggregate level, the OLS estimate of the informal employment earnings gap is a rather huge -25% (Figure 1A; Table A1 in Appendix). Taking into account the (time-invariant) unobserved individual characteristics (UICs) through fixed effect OLS estimation (FEOLS) reduces the earnings penalty significantly, down to -15%. Thus, nearly half of the gap can be explained by UICs, the most productive workers privileging the formal sector. As always, this standard feature does not tell us much about what specific factors are really at play. On the one hand, the innate ability or the 'talent parabola' is commonly stressed in the literature. On the other hand, many other explanations can be put forward. For instance, UICs may have to do with more efficient social networks to get a formal job. However, the remaining -15% gap, once we control for UICs, highlights that formal jobs provide higher earnings *per*

se. Here again, this result can be due to various factors which end up, at the firm level, in a higher productivity or market power and/or, at the worker level, in a stronger bargaining power of formal workers to negotiate higher earnings.

To go beyond average, we ran quantile regressions (QR, Figure 1A; Tables A4 and A5 in Appendix). While suffering earnings penalties at all levels of the conditional distribution, informal workers suffer as smaller gap at the bottom part. From around -23% for the first two quartiles of income, the gap increases to reach -30% at the upper-tier of the distribution (quantile .90). However, the bootstrapped 95% confidence intervals are too large for the estimated QR gaps to be significantly different from the OLS estimator. The Fixed Effects Quantile Regression (FEQR) gap not only confirms the key role of UICs in reducing the 'true' gap but also reveals a remarkable change in the pattern along the earnings distribution. Opposite to the estimated QR gaps, the FEQR gaps are decreasing continuously along the earnings distribution, from19% for the bottom quantile to 10% for the upper one. If a higher number of observations could have increased the precision of our estimates, this result is mainly due to the fact that the 'dualistic assumption' is too rough, gathering together very diverse categories of workers within each sector. As discussed earlier, the informal employment is highly heterogeneous.

Formal vs. informal wage workers

As expected, within wage workers, those employed informally are on average worse-off than those formally hired, the global picture being similar to the one observed for all workers (Figure 1B; Table A1, column (3)). The -23% OLS gap is consistent with the ones observed by Rand and Torm (2012a) on their specific 2009 micro-firms survey (-10% to -20%). It is significantly reduced to -11% when individual fixed effects are introduced, suggesting that informal wage workers may have a disadvantage in terms of their unobserved productive attributes. Not taking into account the fixed effects, no clear distributional effects can be identified; which is not the case when controlling for UIC (Figure 1B; Tables A4 and A5 in Appendix): the gap is continuously decreasing from -16% (quantile .10) to -5% (quantile .90). Nevertheless, in both cases, formal salaried workers conserve an earnings advantage at any position in the pay ladder. Even if we cannot exclude that non-pecuniary disadvantages of formal wage jobs may be compensated by earnings (such as poor working conditions)[9], these results could be taken as an acceptable validation

---

[9] Fernández and Nordman (2009) provide suggestive evidence of pecuniary compensations for working conditions varying along the earnings distribution in the case of the UK.

of the *exclusion hypothesis* (for this category of workers), according to which informal wage workers are constraint in their job choice, and are probably queuing for formal jobs.

Formal wage vs. informal self-employed workers

For the bulk of the labour force, this alternative choice is probably the main trade-off, and also the most discussed in the literature. At odd with the previous case considered and more generally the dualistic approach, the conditional OLS gap is positive, with a significant premium of +6% for the informal self-employed (Figure 1C; Table A1, column (3)). Furthermore, the FEOLS models increase the premium further to +14% (column (5)). This would mean that informal self-employed workers have a disadvantage in terms of their unobserved productive characteristics (probably entrepreneurial skills), which produces an underestimation of the premium associated with being an informal self-employed worker compared to exerting as a formal wage worker if this individual heterogeneity is not accounted for. We nevertheless should be cautious before claiming that the *exit option* may be at stake, as the self-employed earnings may be overestimated for at least two reasons: first, the measure of earnings we computed remunerates both labour and physical capital factors, the latter being far from negligible in the informal sector; second, the self-employed earnings include the share which should be attributed to the productive contribution of unpaid family workers. As we do not have any order of magnitude of these two phenomena, it is difficult to exclude the possibility that the premium we obtain may not turn into a penalty, once these two factors are taken into account.[10]

When turning to quantile regressions (Figure 1C; Tables A4 and A5 in Appendix), the distributional profile of the gap presents a clear pattern, contrary to that of the formal vs. informal wage workers. The gap steeply increases with earnings level, and is in favour of the informal self-employed workers. In absolute terms, informal self-employed labourers suffer a penalty only at the lowest end of the conditional distribution (up to about quantile .30). Afterwards, the gap is reversed into a significant premium, growing continuously up to around 35% for the richest decile, crossing the OLS estimate at the median point of the earnings distribution. FEQRs confirm this trend, the only difference being that the range of variation of the gap along the distribution is attenuated. However, once the UICs are controlled for, informal self-employed workers are better-off at all points of the pay scale, from +2% at quantile .10 to +28% at quantile .90.

---

[10] The definitive assessment is even more complex as measurement errors in incomes are usually considered as more important for self-employed than for wage workers, as the former usually do not know their precise level of income (especially informal workers who do not have book accounts), and the richest ones tend to understate their level of activity.

All in all, keeping in mind the abovementioned limitations of this comparison, and given the size of the estimated premium, we are confident with the finding that informal self-employment may be more lucrative that formal wage alternatives, especially for the richest workers. As a matter of consequence, we have good presumptions to assert that, in Vietnam, a substantial part of the labour force has deliberately chosen to work in the informal sector as non-wage workers, for pecuniary reasons.

Formal vs. informal self-employed workers

We turn to the comparison between the two kinds of self-employed workers: formal and informal. Formal self-employed workers are rarely considered in the literature, maybe because they are too few in the countries considered. That is clearly not the case in Vietnam, as they represent 13% of our sample and more than 40% of the self-employed workers. Furthermore, there are many additional reasons to focus on this category of workers: first, to compare our results with those obtained in developed countries on salaried vs. non-salaried workers' earnings gap, as in these countries self-employed workers are almost exclusively formal; second, because it allows us to establish the link with the existing formal/informal sector literature from a business perspective (not job). Finally, the comparison appears more legitimate as the nature of incomes and unobserved characteristics potentially at play are in both cases equivalent (which is not true concerning wage workers).

Formal self-employed workers are systematically in a better position than their informal counterparts, all along the pay scale (Figure 1D; the reference group is now informal self-employed workers; regressions not shown). Returns to firm's formalisation is always positive and increasing with the net earnings, even when controlling for entrepreneurial skills and other unobserved characteristics, the most favoured in this respect choosing disproportionately the formal sector. This advantage of formal household businesses may be due to higher initial level of physical capital or more productive combination of factors (our models do not provide elements on this point), but it is compatible with the potential intrinsic benefits of getting formal (access to credit and markets) as found by Rand and Torm (2012b) in the case of Vietnam and McKenzie and Sakho (2010) in the case of Bolivia.

Certainly, endogeneity issues may arise if determining factors or time-varying unobserved factors influence both the self-employed workers' decision to formalize and the subsequent business performance. With our data and methodology, in particular the quantile regression

framework, we cannot tackle these issues easily. Yet, using a panel of SMEs surveyed in 2007 and 2009 in Vietnam, Rand and Torm (2012b) show results of the same magnitude using both fixed effects, matched double difference approaches and an IV strategy to account for the fact that firms may self-select into formality. They notably find that firms that became formal during the period 2007-2009 have a 16% higher gross profit growth than comparable firms that remained informal, which is in line with the magnitude of our effect at the mean, that is just below 20%. All in all, most of their estimates suggest that formalization has a positive and well-determined effect on the entrepreneurs' profit. Moreover, as suggested by Rand and Torm (2012b) themselves, their reported formalization effect may be a lower bound estimate due, first, to possible under-reported profits for tax reasons and, second, to their survey's over-representation of relatively more competitive (and profitable) informal firms.

Formal wage vs. formal self-employed workers

Lastly, the earnings comparison of formal wage workers and formal self-employed workers is clearly in favour of the latter, whatever the model chosen (Figure 1E; Tables A1, columns (3) and (5)). The OLS estimate presents a +40% premium, just slightly reduced with fixed effects (+32%). Compared to the informal self-employed workers, their unobserved productive attributes may be better than those of the formal wage workers. As in the case of informal self-employed workers, the premium is continuously increasing with earnings levels, but is translated upwards, a pattern in line with the empirical results obtained in the literature for developed countries. Controlling for UICs or not, formal self-employed workers are always better-off in terms of earnings than formal wage workers, the premium culminating at +70% (QR) or +47% (FEQR). Overall, it seems that the Vietnamese labour market functions under a regime of wage repression. Whatever the reasons - macro pressures of international integration or deliberate policies to control inflation, or weak bargaining power of the wage workers -, it seems globally preferable to work as an independent (even in the informal sector) than as a wage worker (at least in non-farm activities).

A gender perspective

Exploring the gender dimension associated with informality is crucial for various reasons. First, there is strong imbalances in the job structure, female being more prone to hold informal jobs than their male counterparts. Second, the raw gender earnings gap is in general

significantly higher in the informal sector.[11] Finally, and more importantly, the motivation to hold informal jobs is highly dependent on gender. Women may have a welfare function which is less dependent on income incentives, as they take more care of extra professional activities, where informal jobs could be a more satisfying option. We highlight below the main findings displayed in Figures 2 and Figures 3 in Appendix and their corresponding regression Tables A2, A3 and A6 to A9.

Firstly, whatever the model specifications and the category of workers considered, females always financially suffer more (or benefit less) when they are informally employed. For instance, at the aggregate formal/informal level (Figures 2A and 3A), the OLS gap is -19% for men and -30% for women; the FEOLS being respectively -11% and -20%. Such a feature is compatible with the idea that women may accept lower wages in the informal sector because it provides other non-pecuniary advantages, relatively more valuable to them. However, it can also reveal barriers or labour market segmentation, which would be more pronounced for women competing for salaried jobs. Interestingly, while the penalty for being informal wage workers remains substantial for women once UICs are controlled for (-18%, Figure 3B and column (5) of Table A3), it is no more significant for men. For the latter, working informally is at least financially as rewarding as having a formal job, whether dependent (Figure 3B) or independent (Figure 3C).

Secondly, in spite of differences in absolute levels, the distributional profile of the earnings gaps is quite similar across gender: no noticeable effect for wage workers, an increasing premium for self-employed workers, both formal and informal. The only exception is for informal wage workers, where this type of jobs seems to be relatively more attractive for women in the upper tier, while the gap increases smoothly for men.

Thirdly, the sorting process in the allocation of men and women across employment status (which is partly revealed by the effect of controlling for UICs) does not differ substantially across gender: informal wage workers have detrimental UICs (in order to get a better income) vis-à-vis formal wage workers, while the unobserved skills are favourable for self-employed workers (whether formal or informal). The only exception is for male wage workers, who have comparable UICs along the formal/informal divide.

---

[11] For Africa, Nordman, Robilliard and Roubaud (2011) estimate the gender earnings gap in the formal and informal sectors of different West African capital cities using household surveys and find higher gender gaps for the category of informal self-employed workers.

The Vietnamese case in perspective

Before comparing our results with those obtained in other developing countries, let us briefly remind some important changes on the Vietnam's labour market during the period of the surveys (see also Section 1). This period first witnessed a process of deep domestic and external reforms aiming at accelerating international economic integration. These reforms had large impacts on the relative position of different economic sectors. For instance, the passage of the Enterprise Laws of 2000 and 2005 created good conditions for the development of non-state enterprises, providing substantial employment opportunities outside the state sector. Second, in the context of rapid trade liberalization, the labour market has known significant changes in wage levels, skill premiums and earnings inequalities (Cling *et al.*, 2010b). Previous studies on Vietnam found strong evidence of large wage differentials and labour market segmentation in different dimensions, such as region, gender, and ethnicity (for a detailed literature review, see Phan, 2009). However, most of these empirical studies relied on VLSS 1993 and 1998 data, thereby the available results reflected wage gap trends in the 1990s and early 2000s. A more recent analysis provided in Phan (2009) for the period of the 2000s shows a reversed trend of wage inequality (measured by the Gini index), as well as of the gender wage gap. By contrast, the rural-urban and ethnic wage gaps appear to have increased in the same period.

Comparing our results with those obtained in other developing countries may allow us to highlight the Vietnam's labour market specificities. One point should be stressed however: the number of countries of comparison is rather limited, and is mainly restricted to Latin America; more, to our knowledge, the only paper to undertake FEQR is Bargain and Kwenda (2011), with the limitation that these authors present estimates for a subsample of full time male workers (but with the advantage of much larger sample sizes).

In spite of the unique nature of the Vietnamese economy, our results are in line with the literature, emphasizing the dual nature of informal jobs. Furthermore, as our estimations are also conducted for women, with globally similar gap structures, these converging results may be generalised as a stylized feature. While, on the whole, informal workers suffer penalties vis-à-vis formal workers, this feature is mainly due to informal wage earners. In fact, informal self-employed workers receive increasing premium vis-à-vis formal wage workers along the pay ladder. This feature put Vietnam closer to Mexico (and to a lesser extend Brazil) than to South Africa, where the gap, although decreasing, is always negative, even at the highest end of the earnings distribution. Vietnam, although a much poorer country, already exhibits a more integrated labour market, which is a characteristic of

emerging Latin American countries compared to the dualistic Sub-Saharan African countries (see Bocquier *et al.*, 2010). It seems that the labour market segmentation is even less pronounced there than in the former countries, as it is the only of the four abovementioned countries where informal wage workers do not suffer penalties compared to their formal counterparts (only for males).

Finally, formal self-employed workers represent a very specific job segment, at the top of the earnings hierarchy, which should neither be aggregated with formal wage earners, nor discarded from the analysis as in many other studies.

5. Conclusion

In this paper, we study which of the exclusion or the exit hypothesis regarding informality is best suited to the Vietnamese labour market. To this end, we focus on the earnings gaps between formal and informal workers. Assuming that individual earnings are proxies of individual utilities, our approach considers that if informal workers earn more than their formal counterparts, this reflects a deliberate choice of the former to be informal workers. Taking advantage of the VHLSS three wave panel dataset (2002-2004-2006), we address the key issue of heterogeneity at the worker, the job, and the distributional levels. Our results suggest that the informal earnings gap highly depends on the workers' job status and on their relative position in the earnings distribution. Our main conclusion is at odds with the exclusion hypothesis and what would show the observed raw earnings gaps: in many cases, informal jobs are more rewarding (self-employment) or as rewarding (male wage workers) as formal wage jobs. This feature is due to the relatively low wages of formal wage jobs. The reason for such a specificity should be investigated further: International competition pressure? Wage repression policy? Second, Vietnam's labour market seems more integrated than what its development level would have predicted. The earnings gaps look more like those observed in emerging countries, characterised by a weak segmentation between formal and informal jobs, than the standard dualistic Sub-Saharan labour markets. Third, the systematic premium at all points of the distribution of formal self-employed workers over their informal counterparts suggests that formalization of non-farm household businesses seems to be beneficial, thereby confirming previous studies on this issue. Policies aiming at easing administrative procedures to register informal firms should be encouraged. Finally, females always financially suffer more (or benefit less) when they are informally employed. This feature opens space for specific policies to align the functioning of labour market for

women with that of men (reduction in entry barriers to formal jobs, improvement of access to physical capital, etc.).

Our paper raises further promising prospects, and could be extended in various directions. A first extension would be to better control for individual unobserved characteristics, by purging our earnings estimations of differences in the amount of physical capital (for self-employed workers) and social networks. A firm-based panel approach is an interesting alternative entry in this respect (see Nguyen and Nordman, 2012 using the VHLSS data). Another potential extension would be to exploit further the nature of our data by estimating dynamic earnings equations. Lastly, our work could be usefully complemented by investigating the determinants of job satisfaction, to enlarge the perspective which relies exclusively on earnings outputs and to check the robustness of our conclusions in this regard.

References

Bargain, O., Kwenda, P. (2011), "Earnings Structures, Informal Employment, and Self-Employment: New Evidence from Brazil, Mexico, and South Africa", *Review of Income and Wealth*, 57, Special Issue, pp. S100-S122.

Belsley, D.A., Kuh, E., Welsch, R.E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley, New York.

Bocquier, P., Nordman, C.J., Vescovo, A. (2010), "Employment Vulnerability and Earnings in Urban West Africa", *World Development*, 38(9), pp. 1297-1314.

Canay, I.A. (2011), "A Simple Approach to Quantile Regression for Panel Data", *The Econometrics Journal*, 14(3), pp. 368-386.

Cling, J.-P., Marouani, M.A., Razafindrakoto, M., Robilliard, A.-S., Roubaud, F. (2009), "The Distributive Impact of Vietnam's Accession to the WTO", *International Economics*, No. 118, pp.45-73.

Cling, J.-P., Nguyễn, ThịThu Huyền, Nguyễn, Hữu Chí, Phan, T. Ngọc Trâm, Razafindrakoto, M., Roubaud, F. (2010a), *The Informal Sector in Vietnam: A Focus on Hanoi and Ho Chi Minh City*, Hanoi: The Gioi Edition.

Cling, J.-P., Nguyễn, Hữu Chí, Razafindrakoto, M., Roubaud, F. (2010b), "*Urbanization and Access to Labour Market in Vietnam: Weight and Characteristics of the Informal Sector*", in Ton, Nu Quynh Tran, Quertamp F., de Miras, C., Nguyen, Quang Vinh, (eds.), *Trends of Urbanization and Periurbanization in South East Asia,* Ho Chi Minh City: HCMC Publishing house, AFD, PADDI, IMV.

Cling, J.-P., Razafindrakoto, M., Roubaud, F. (2012), "To Be or Not To Be Registered? Explanatory Factors Behind Formalizing Non-farm Household Businesses in Vietnam", *Journal of the Asia Pacific Economy*, 17(4), pp. 632-652.

De Soto, H. (1989), *The Other Path: The Invisible Revolution in the Third World*, New York: Harper and Row.

Falco, P., Kerr, A., Rankin, N., Sandefur, J., Teal, F. (2010), "The Returns to Formality and Informality in Urban Africa", CSAE WPS 2010-03, Oxford.

Fernández, R., Nordman, C.J. (2009), "Are There Pecuniary Compensations for Working Conditions?", *Labour Economics*, 16(2), pp. 194-207.

Fields, G.S. (1990), "Labour Market Modelling and the Urban Informal Sector: Theory and Evidence", In *The informal sector revisited*, Turnham D., Salomé B., & Schwarz A. (Eds.), Paris: Organisation for Economic Co-operation and Development, pp. 49-69.

Gong, X., Van Soest, A., Villagomez, E. (2004), "Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico", *Economic Development and Cultural Change*, 53(1), pp. 1-36.

Harris, J.R., Todaro, M.P. (1970), "Migration, Unemployment, and Development: A Two-sector Analysis", *American Economic Review*, 60(1), pp. 126-42.

ILO (2003), *Guidelines Concerning a Statistical Definition of Informal Employment*, Seventeenth International Conference of Labour Statisticians, ILO, Geneva.

Imbert, C. (2011), "Decomposing Wage Inequality: Public and Private Sectors in Vietnam 1993-2006", Paris School of Economics, Working Paper No.2011-05, Paris, January.

Koenker, R. (2004), "Quantile Regression for Longitudinal Data", *Journal of Multivariate Analysis*, 91, pp. 74-89.

Lewis, W.A. (1954), "Economic Development with Unlimited Supplies of Labour", *Manchester School*, 28(2), pp. 139-191.

Maloney, W. (1999), "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico", *World Bank Economic Review*, 13(2), pp. 275-302.

Maloney, W. (2004), "Informality Revisited", *World Development*, 32(7), pp. 1159-1178.

McCaig, B., Benjamin, D., Brandt, L. (2009), "The Evolution of Income Inequality in Vietnam, 1993-2006", mimeo, Australian National University and University of Toronto.

McKenzie, D., Sakho Y.S. (2010), "Does it Pay Firms to Register for Taxes? The Impact of Formality on Firm Profitability", *Journal of Development Economics*, 91(1), pp. 15-24.

Migheli, M. (2012), "Do the Vietnamese Support the Economic *Doi Moi*?", *Journal of Development Studies*, 48(7), in press.

Nguyen, H.C., Nordman, C.J. (2012), "Household Entrepreneurship and Social Networks: Panel Data Evidence from Vietnam", Unpublished Manuscript, DIAL, Paris.

Nordman, C.J., Robilliard, A.-S., Roubaud, F. (2011), "Gender and Ethnic Earnings Gaps in Seven West African Cities", *Labour Economics*, Volume 18, Supplement 1, pp. S132-S145.

Oostendorp, R.H., Tran, Quoc Trung, Nguyen, Thanh Tung (2009), "The Changing Role of Non-Farm Household Enterprises in Vietnam", *World Development*, 37(3), pp. 632-644.

Perry, G.E., Maloney, W.F., Arias, O.S., Fajnzylber, P., Mason, A.D., Saavedra-Chanduvi, J. (2007), *Informality: Exit and Exclusion*, Washington DC: The World Bank, World Bank Latin American and Caribbean Studies.

Phan, D. N. (2009) "A Report on Vietnam's Labour Market", CIEM-DANIDA Project: Poverty Reduction Grant.

Rand, J., Torm, N. (2012a), "The Informal Sector Wage Gap Among Vietnamese Micro-firms", *Journal of the Asia Pacific Economy*, 17(4), pp. 560-577.

Rand, J., Torm, N. (2012b), "The Benefits of Formalization: Evidence from Vietnamese SMEs", *World Development*, 40(5), pp. 983-998.

Razafindrakoto, M., Roubaud, F., Wachsberger, J.-M. (2012), "Travailler dans le secteur informel : choix ou contrainte ? Une analyse de la satisfaction dans l'emploi au Vietnam", In Cling, J.-P., Lagrée, S., Razafindrakoto, M., Roubaud, F. (eds.), *L'économie informelle dans les pays en développement*, Paris: AFD Editions, Chapter 1.2, pp. 47-66.

Roubaud, F. (1994), *L'économie informelle au Mexique : de la sphère domestique à la dynamique macro-économique*, Paris: Karthala/Orstom.

Söderbom, M., Teal, F., Wambugu, A. (2005), "Unobserved Heterogeneity and the Relation Between Earnings and Firm Size: Evidence from Two Developing Countries", *Economics Letters*, 87(2), pp. 153-159.

Strobl, E., Thornton, R. (2004), "Do Large Employers Pay More in Developing Countries? The Case of Five African Countries", *Journal of Economic Development*, 29(1), pp. 137-161.

Tran, QuocTrung, Nguyen,Thanh Tung (2008), "Effects of Trade Liberalization on Non-farm Household Enterprises in Vietnam", DEPOCEN Working Paper Series 2008/32, Hanoi.

VASS (2010), *Poverty Reduction in Vietnam: Achievements and Challenges*, Vietnam's Poverty Assessment 2008-2010: Synthesis report, Hanoi.

Vijverberg, W., Hoang,Thi Thanh Huong, Nguyen, Chien Tang, Nguyen, Ngoc Que, Nguyen, The Quan, Phung, Duc Tung, Vu, Thi Kim Mao (2006), "Non-Farm Household Enterprises in Vietnam. A Research Project using data from VHLSS 2004, VHLSS 2000 and AHBS 2003", processed, Hanoi, July.

Appendix

Additional Tables and Figures

****

Box 1. Building the panel of non-farm workers with VHLSS 2002, 2004 and 2006

The construction of the panel is a process of two steps: in the first step, we match different databases from different modules for each year, and then in a second step we match the years. This proves to be a complicated process as there arose some ambiguities in both steps that we summarise below.

	2002	2004	2006
Full sample (household)	75,000	45,000	45,000
Detailed sample (household)	30,000	9,000	9,000
All individuals			
- Unbalanced Panel	18,299	27,828	16,937
- Balanced panel	7,408	7,408	7,408
Population aged 10 years or over*			
- Unbalanced Panel	13,732	23,326	15,336
- Balanced panel	5,742	5,742	5,742
Non-farm workers aged 15 years or over			
- Unbalanced Panel	**2,498**	**4, 395**	**2,845**
- Balanced Panel	**948**	**948**	**948**
- Observed in 2002 and 2004	**1,550**	**1,550**	**-**
- Observed in 2004 and 2006	**-**	**1,897**	**1,897**

Source: VHLSS, 2002, 2004 & 2006, GSO; authors' calculations.
Notes:
*: For which information on employment was asked for in 2002. In 2004 and 2006, this section of questions was applied for the population aged 6 years and over. **: 15 years old as of the start of the panel (2002 or 2004), which corresponds to the minimum age at which a person may be employed at the end of compulsory schooling according to the ILO convention.

Our balanced panel includes 7,408 individuals matched between all the three rounds of VHLSS (see Table above): 10,891 individuals observed only in 2002 and 2004 and 9,529 individuals observed only in 2004 and 2006.

As the major objective of our study is to investigate the question of earnings of workers participating in formal/informal employment in private or household enterprises, we have integrated information of the module on non-farm household businesses into the individual level data. Some difficulties have arisen when we matched the files in 2004. There was no information in the non-farm household business modules of the 2004 VHLSS (M10 and M4C) to identify exactly the 'most knowledgeable' household member to be considered as the head of the household business. As key variables for matching the different modules, we hence used, on one side, the household identifier together with the branch code of jobs of occupied members (available from the module on individual socio-demographics) and, on the other side, the branch code of non-farm businesses of the household. This resulted in uncertainties or non-matched cases since there might have been errors during the coding of branches. To tackle this issue, before matching, we retained household occupied members who were identified in the module of employment as engaged in non-farm household activities as their main job. This helped excluding from each household all the occupied members who were not working in non-farm household activities, whose industry codes of main job resembled that of other non-farm self-employed members.

Box 2. Elaborating regional and time deflators

As the regional deflators (16 locations, that is 8 regions in two areas, urban and rural) included in the VHLSS databases have been criticised for not being consistent over time (McCaig *et al.*, 2009), we combined the VHLSS 2006 regional deflators (supposed to be the most reliable) with the provincial CPIs (63 provinces) provided by the General Statistics Office aggregated at the regional level. We proceed as follows:

Step 1: the 16 official spatial deflators obtained from the VHLSS2006 have been chosen as the reference (1=national price average for 2006);

Step 2: CPIs by regions (8) have been computed by aggregating (simple average) and retropolating to 2004 and 2002 the official monthly provincial GSO CPIs, taking into account the changes in boundaries of regions and provinces over time (1=regional CPIs in 2006). As provincial GSO CPIs do not distinguish the rural and urban divide, the same CPIs have been applied to the two areas over time;

Step 3: step 1 and step 2 have been combined to elaborate consistent regional CPIs, which incorporate both spatial differentials and time dynamics.

Results are shown in the Table below. The adjustment is quite substantial given the high differences in price levels and inflation: a difference of more than 77% in prices is observed between the lowest price level (rural North-East region, 2002) and highest one (urban South-East region, 2006), showing that markets are far from being fully integrated in Vietnam.

	2002		2004		2006	
	Urban	Rural	Urban	Rural	Urban	Rural
Red Delta River	0.918	0.853	1.017	0.945	1.083	1.007
North East	0.737	0.694	0.833	0.785	0.962	0.907
North West	0.760	0.736	0.862	0.835	1.019	0.988
North Central Coast	0.744	0.643	0.846	0.731	0.995	0.861
South Central Coast	0.809	0.736	0.908	0.827	1.072	0.976
Central Highlands	0.808	0.727	0.878	0.789	1.035	0.93
South East	0.929	0.798	1.038	0.892	1.233	1.06
Mekong Delta River	0.843	0.737	0.935	0.818	1.095	0.958

Figure 0. Distribution of Movers in/out of the Informal Employment (%)
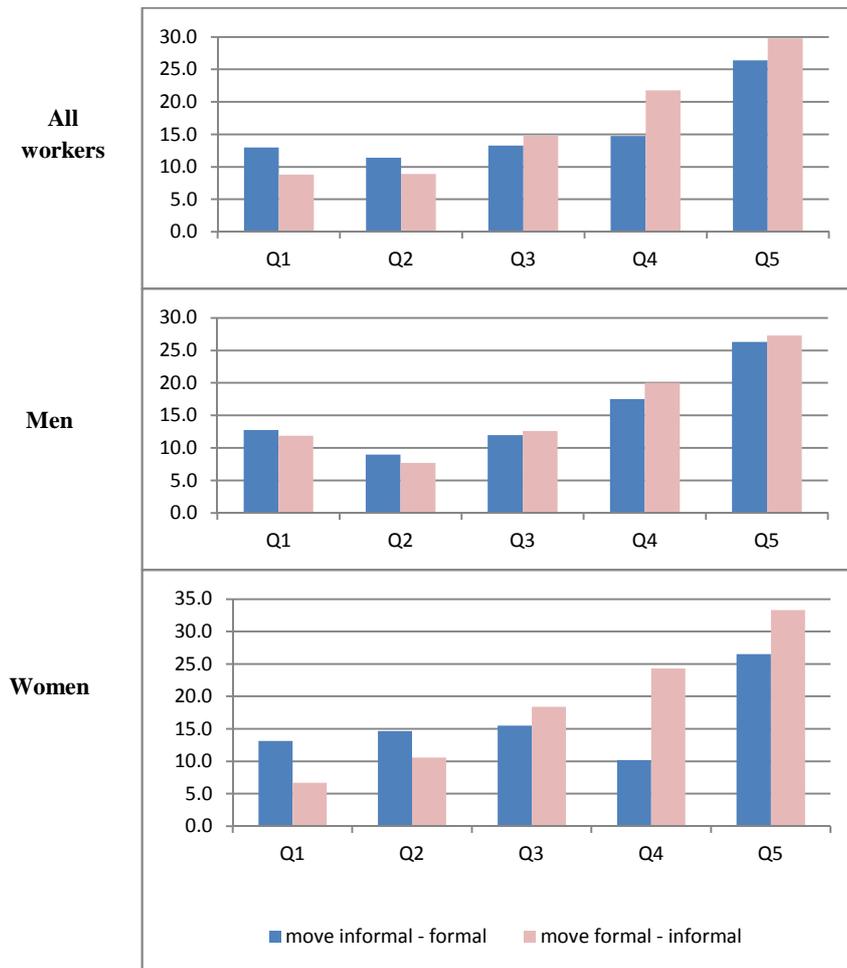
Source: VHLSS, 2002 & 2006, GSO; authors' calculations.
Note: Quintiles of real hourly earnings in 2006.

Figures 1. Estimated Earnings Gaps for Full Sample of Men and Women
(with reference to formal wage workers)

Note: Fixed Effects (FE) OLS are denoted by FEOLS and Fixed Effects Quantile Regressions (QR) by FEQR. Bootstrapped 95% confidence intervals are represented by the grey surface for QR and by dashed lines for the OLS.

1.A Informal Worker -Formal Worker Earnings Gap
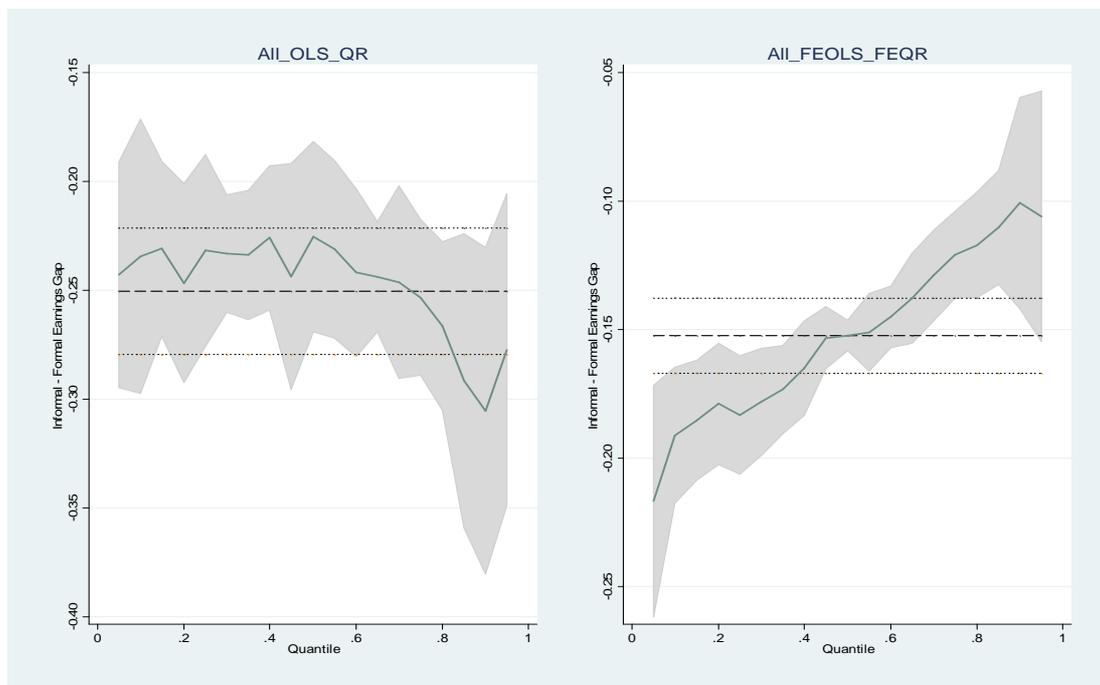
1.B Informal Wage Worker - Formal Wage Worker Earnings Gap
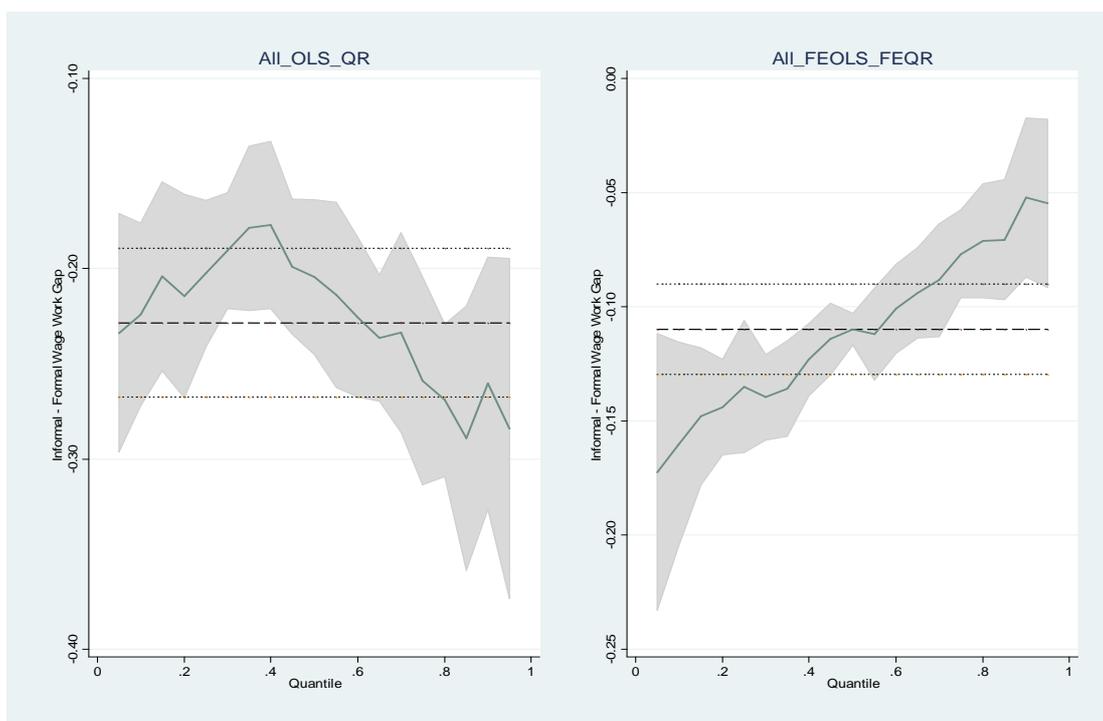
1.C Informal Self-employed Worker - Formal Wage Worker Earnings Gap
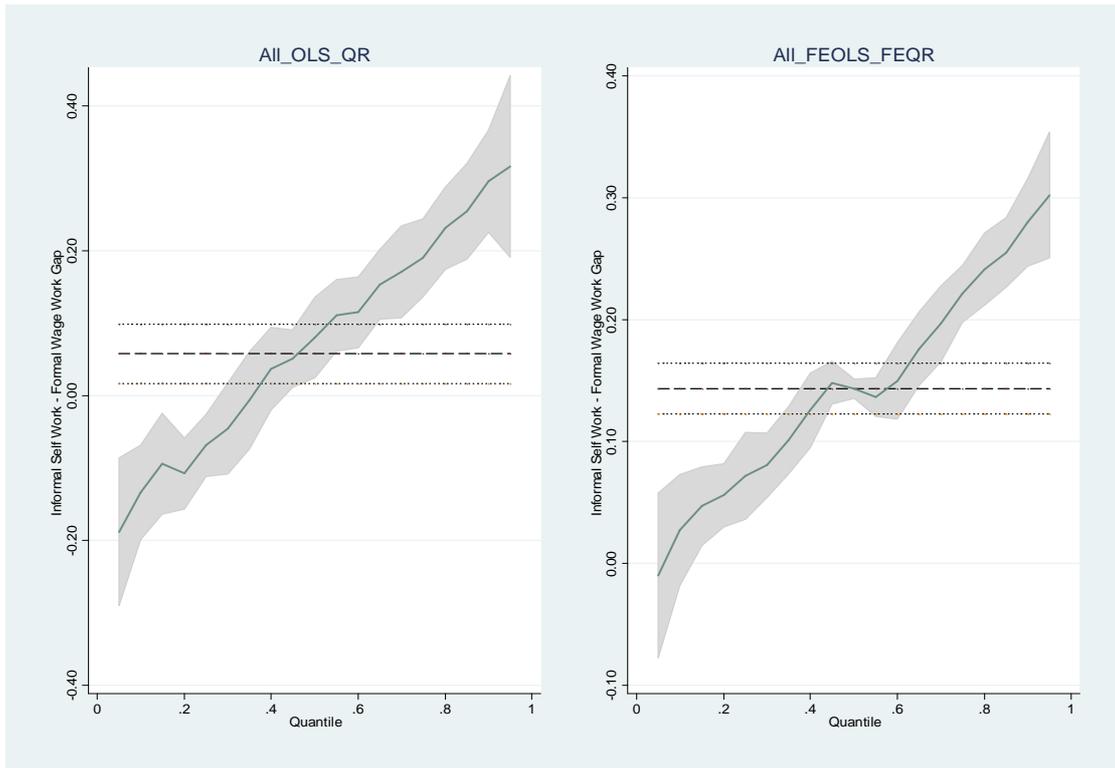
1.D Formal Self-employed Worker - Informal Self-employed Worker Earnings Gap

1.E Formal Self-employed Worker - Formal Wage Worker Earnings Gap

Figures 2. Estimated Earnings Gaps for Women and Men Separately by OLS and QR
(with reference to formal wage workers)

Note: Fixed Effects (FE) OLS are denoted by FEOLS and Fixed Effects Quantile Regressions (QR) by FEQR. Bootstrapped 95% confidence intervals are represented by the grey surface for QR and by dashed lines for the OLS.

2.A Informal Worker -Formal Worker Earnings Gap – OLS QR

2.B Informal Wage Worker - Formal Wage Worker Earnings Gap – OLS QR
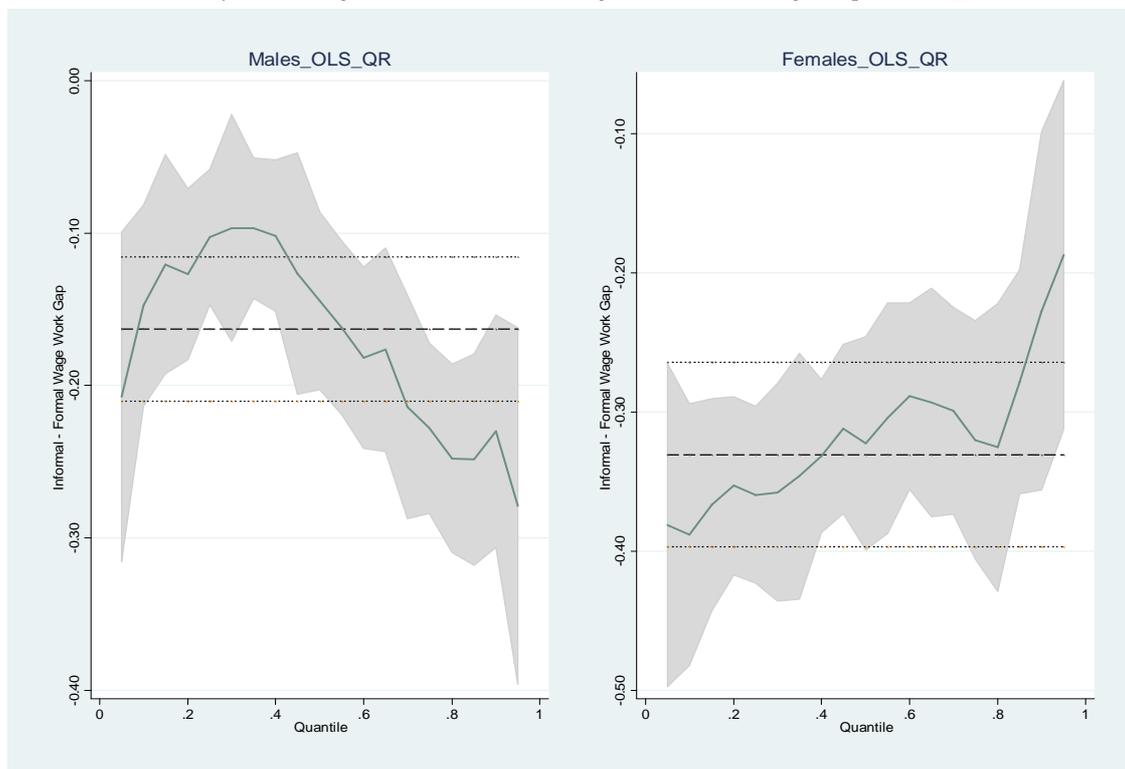
2.C Informal Self-employed Worker - Formal Wage Worker Earnings Gap – OLS QR

2.D Formal Self-employed Worker - Formal Wage Worker Earnings Gap – OLS QR

2.E Formal Self-employed Worker - Informal Self-employed Worker Earnings Gap – OLS QR

Figures 3. Estimated Earnings Gaps for Women and Men Separately by FEOLS and FEQR
(with reference to formal wage workers)

3.A Informal Worker -Formal Worker Earnings Gap – FEOLS FEQR
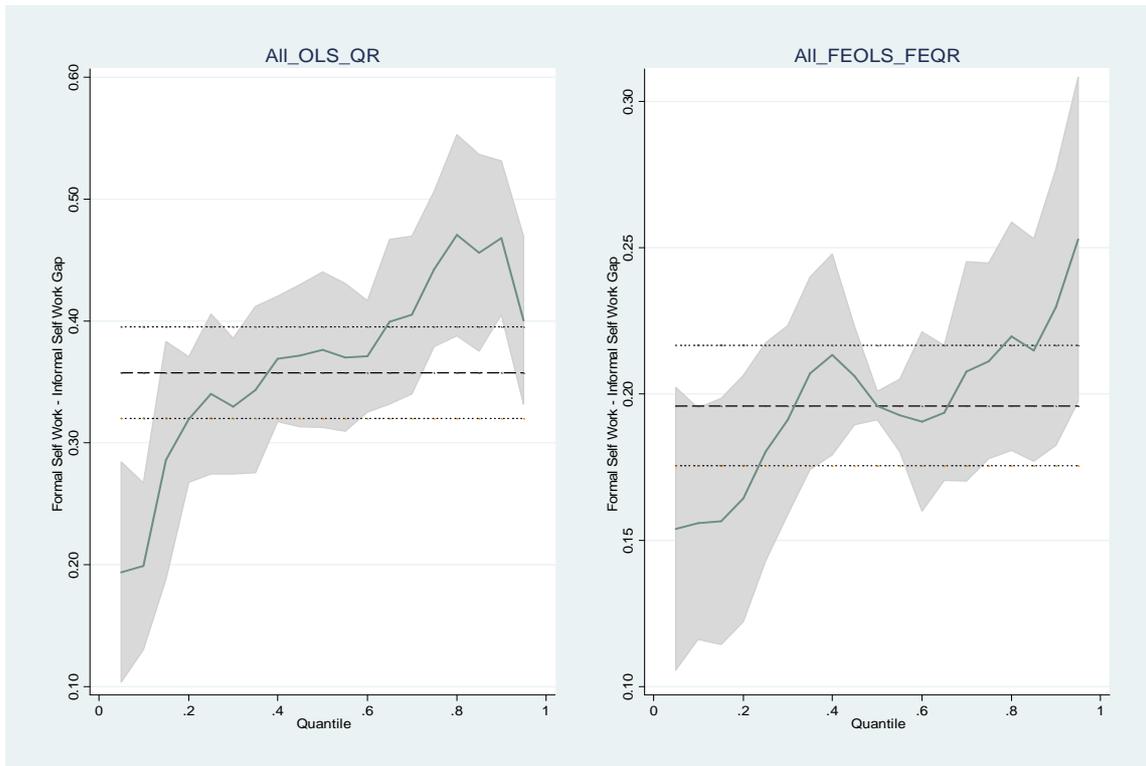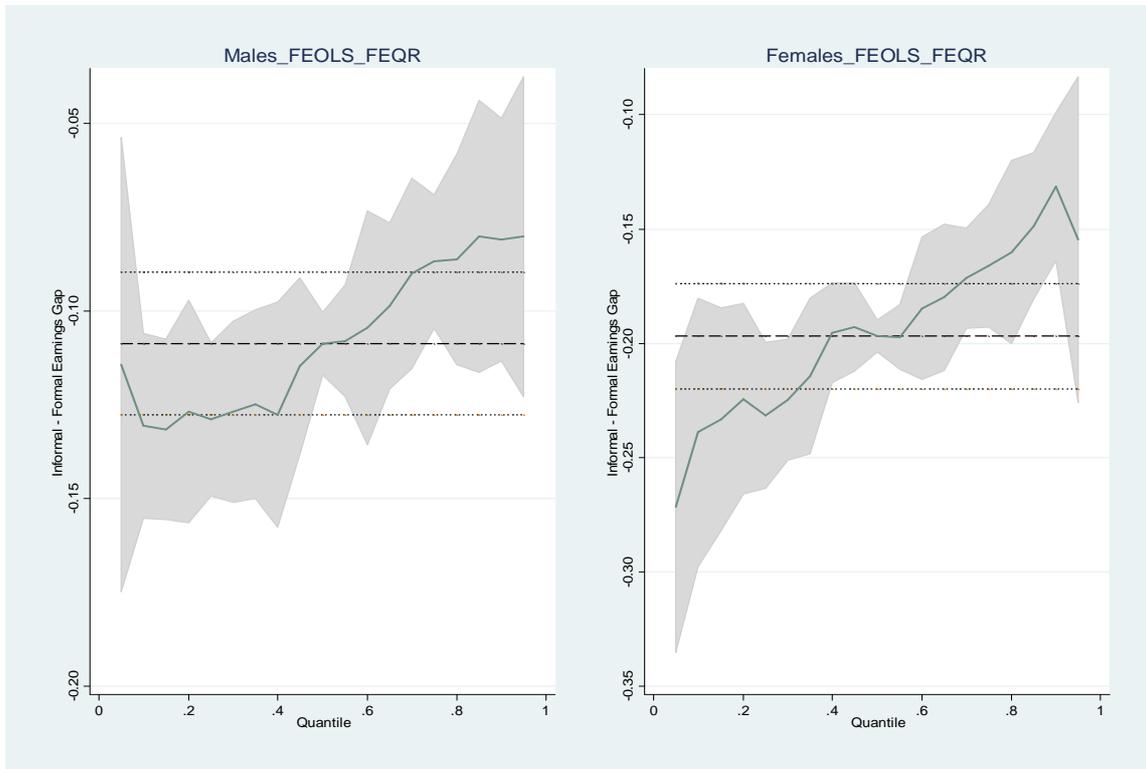

3.B Informal Wage Worker - Formal Wage Worker Earnings Gap – FEOLS FEQR
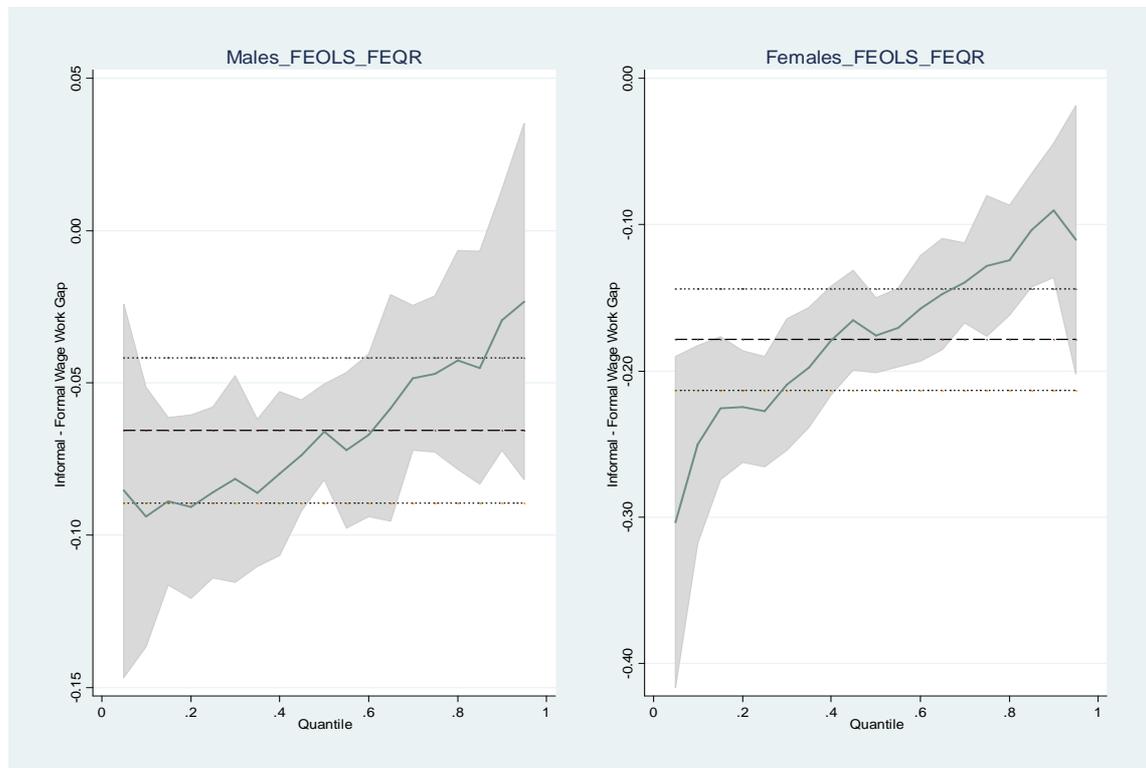
3.C Informal Self-employed Worker - Formal Wage Worker Earnings Gap – FEOLS FEQR
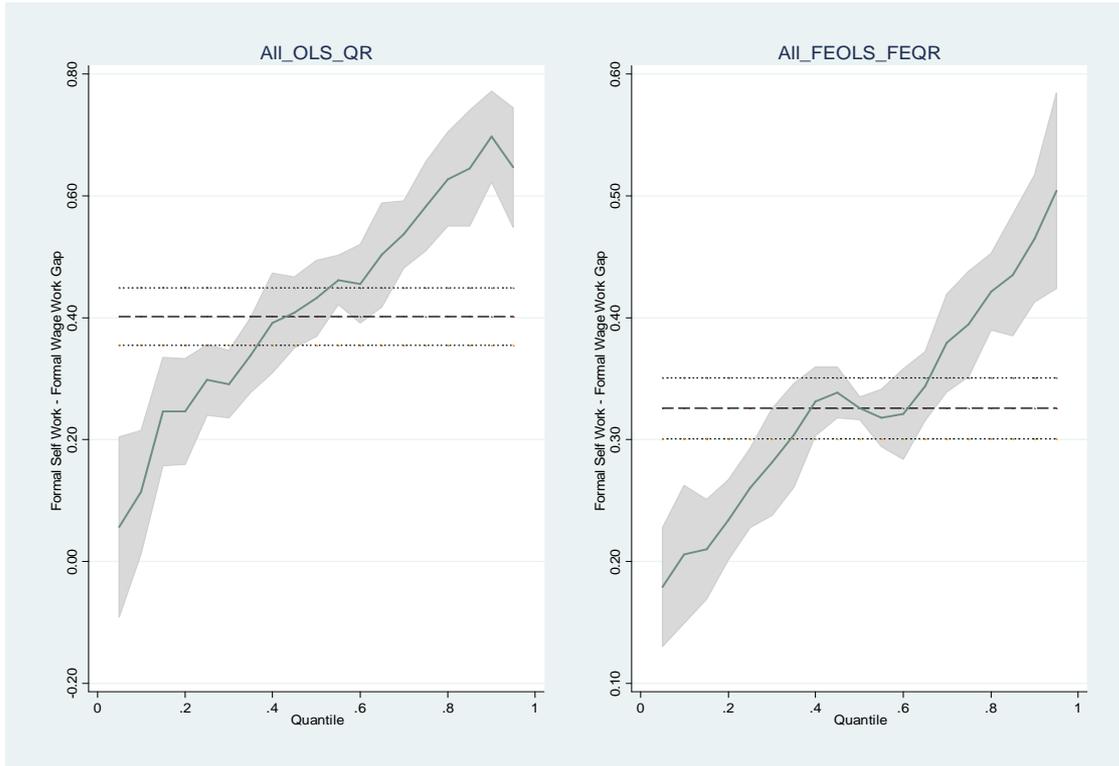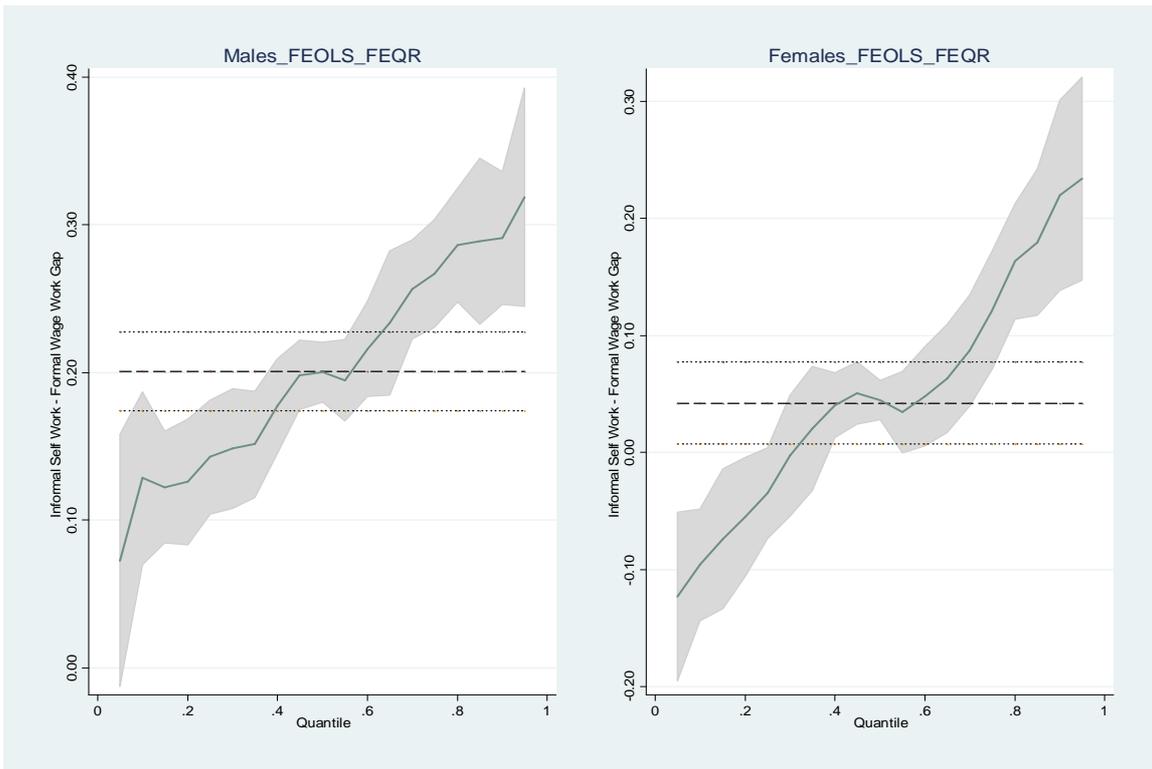

3.D Formal Self-employed Worker - Formal Wage Worker Earnings Gap – FEOLS FEQR

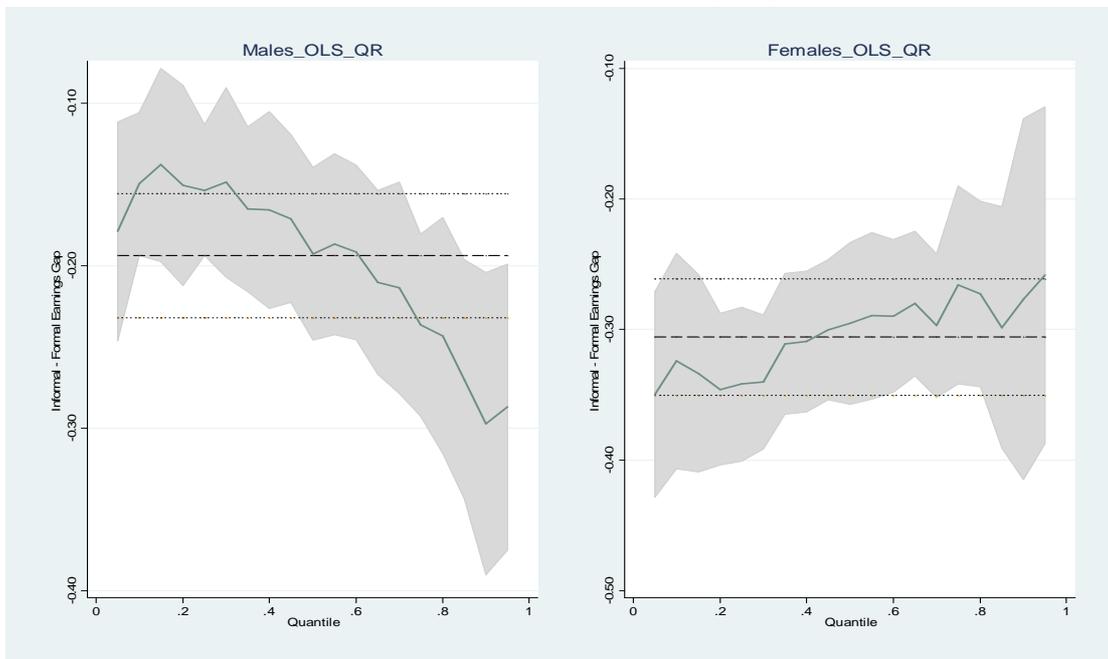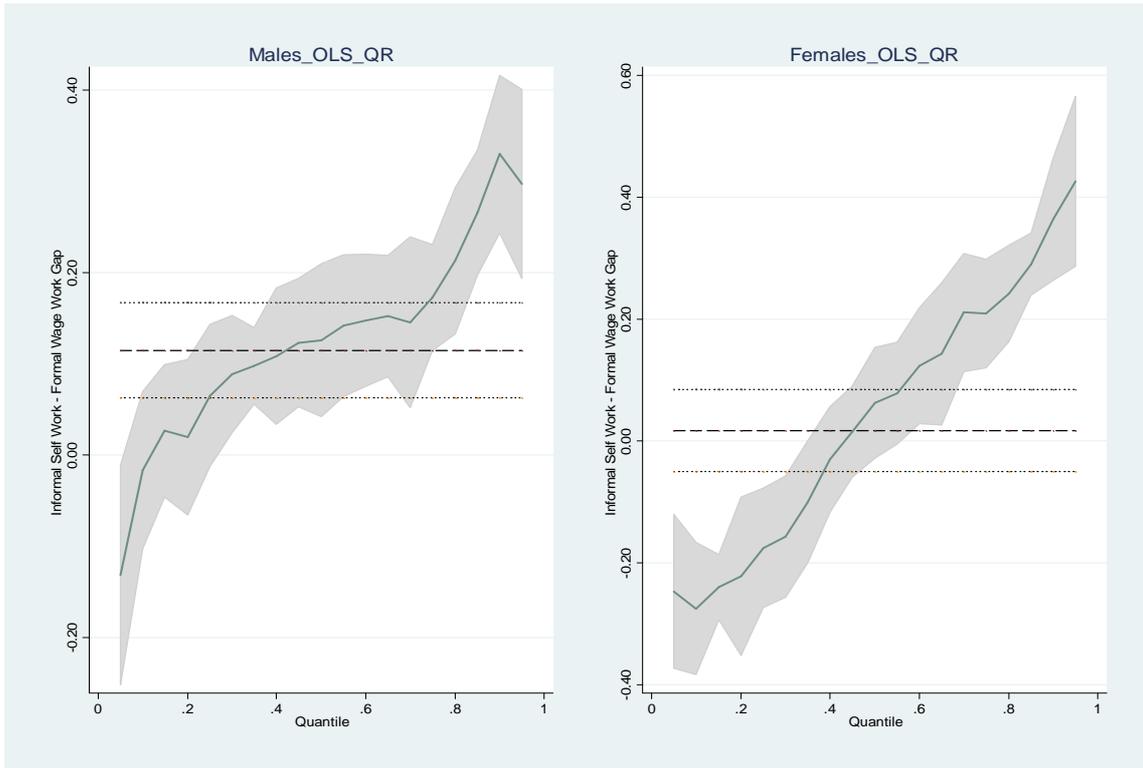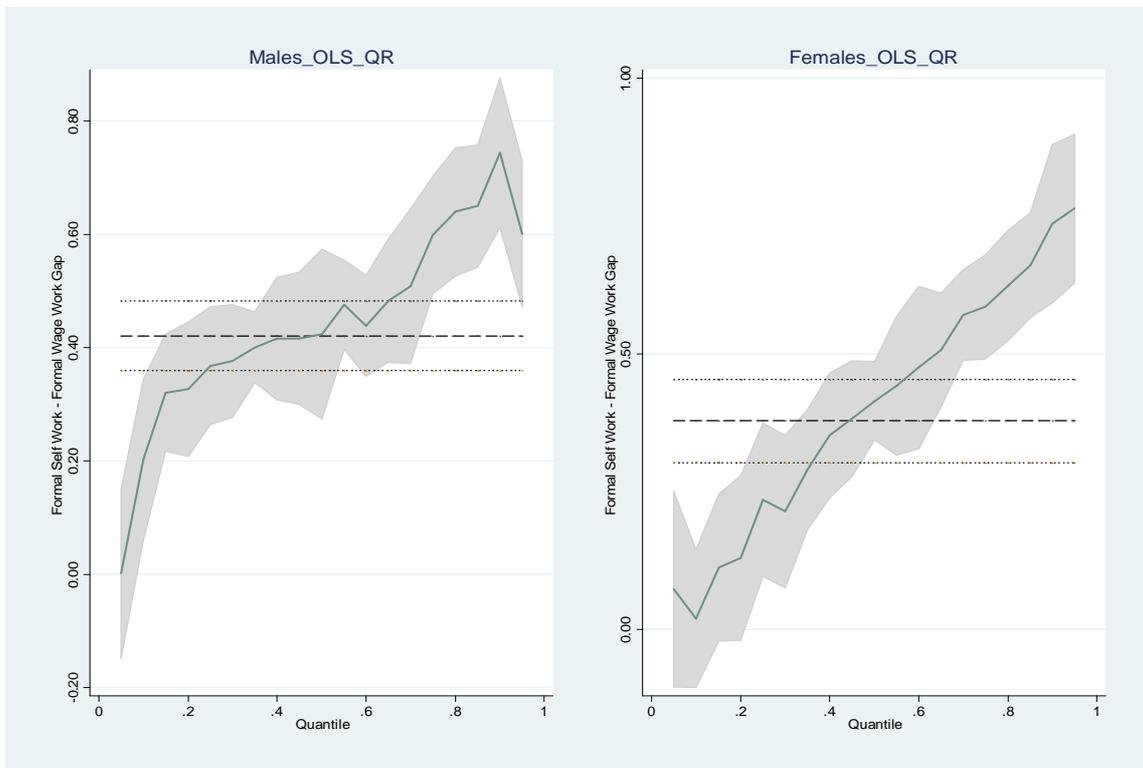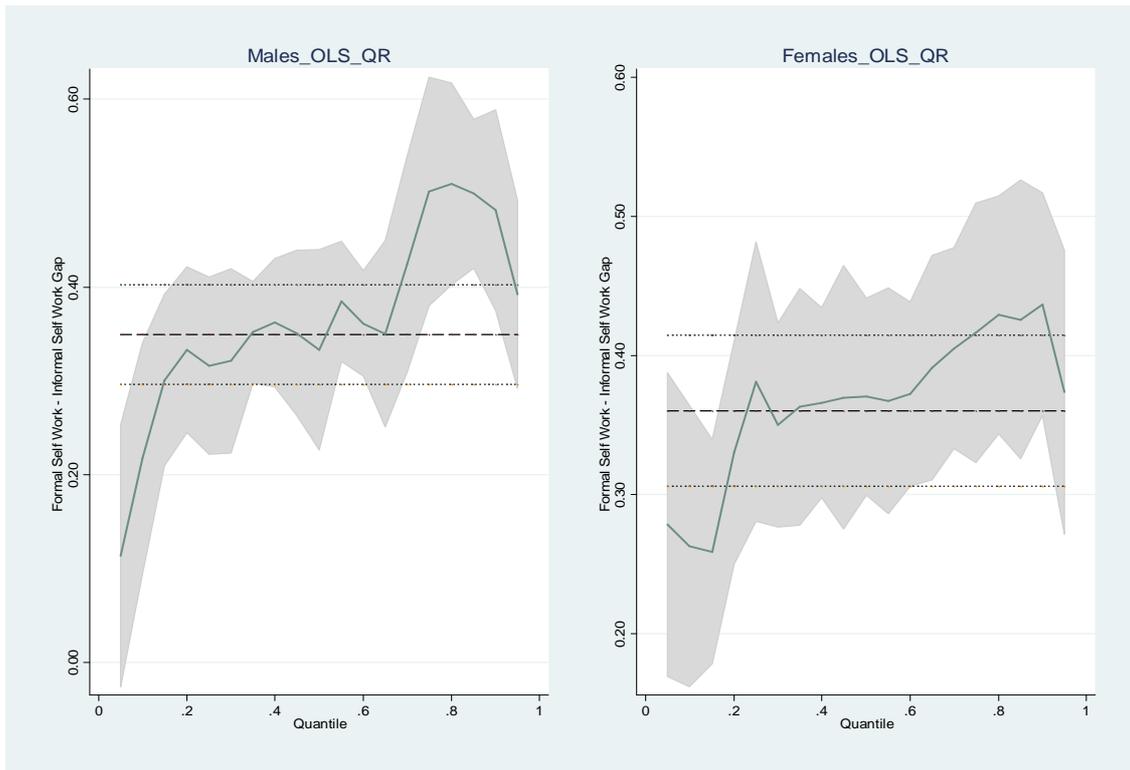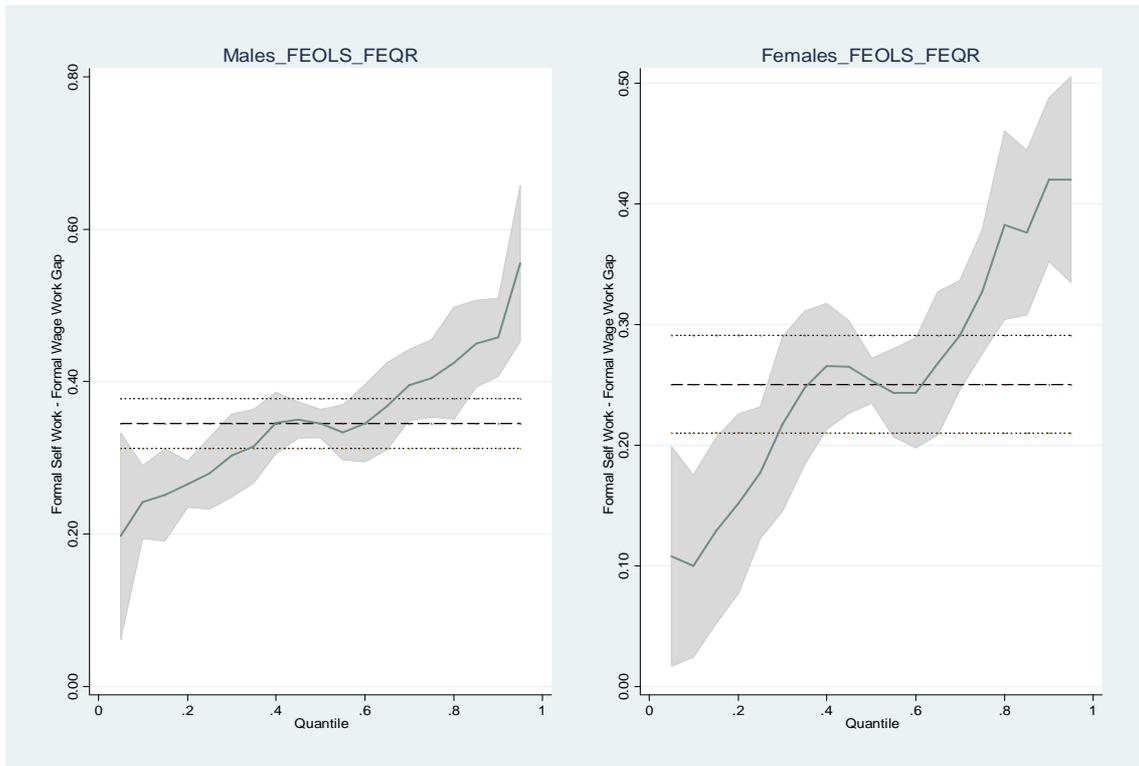*3.E Formal Self-employed Worker - Informal Self-employed Worker Earnings Gap – FEOLS FEQR*
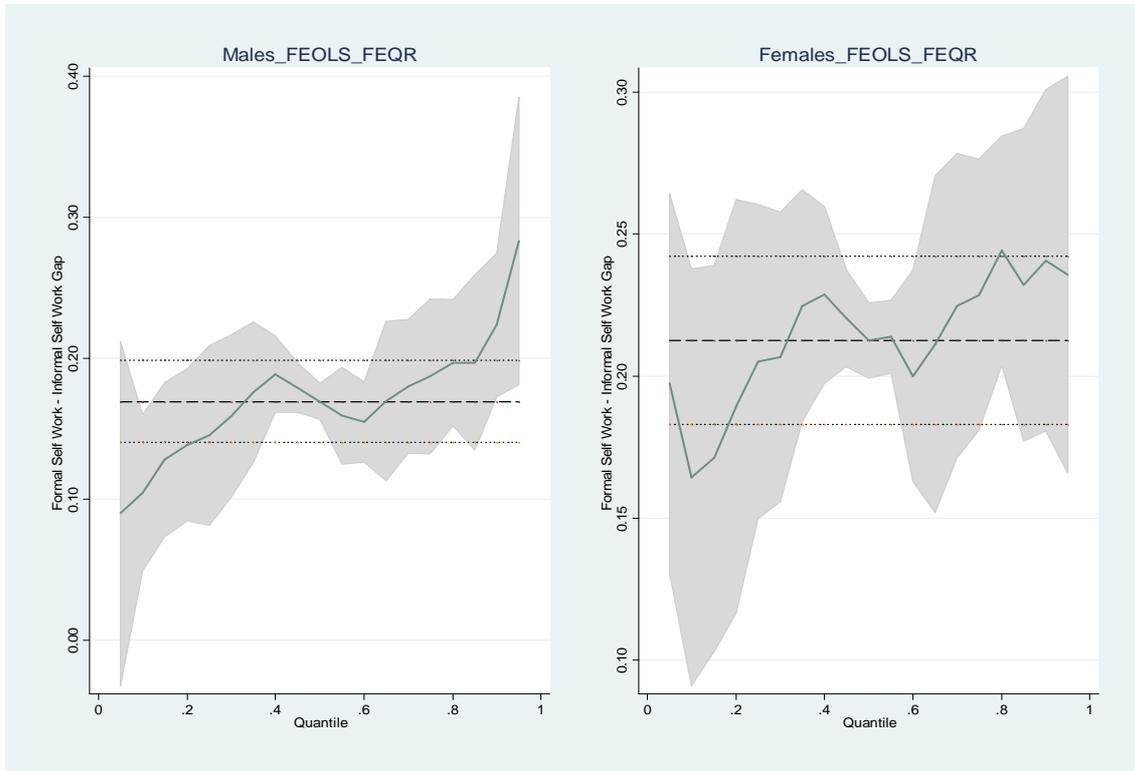
Table A0. Inter-sector Switches and Job Changes (%)

Job changes/Type of transition	2002 – 2004	2004 – 2006
Informal – formal		
Change in occupation	41.9	46.7
Change in industry	36.5	42.7
Change in occupation and/or industry	50.0	62.7
Formal – informal		
Change in occupation	42.4	44.4
Change in industry	40.0	41.7
Change in occupation and/or industry	69.0	59.7

Source: VHLSS, 2002 & 2006, GSO; authors' calculations.

Table A1. Mean Earnings Regressions For All Workers

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled OLS	(2) Pooled OLS	(3) Pooled OLS	(4) Fixed Effects	(5) Fixed Effects
Informal Worker		-0.250***		-0.152***	
		(0.016)		(0.024)	
Informal Self-Emp. Worker			0.058***		0.143***
			(0.021)		(0.045)
Informal Wage Worker			-0.228***		-0.110***
			(0.018)		(0.030)
Formal Self-Emp. Worker			0.402***		0.326***
			(0.027)		(0.053)
Years of schooling	-0.015**	-0.017***	-0.033***		
	(0.006)	(0.006)	(0.006)		
Years of schooling squared	0.004***	0.004***	0.005***		
	(0.000)	(0.000)	(0.000)		
Potential experience	0.027***	0.025***	0.023***	0.027***	0.026***
	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)
Potential experience squared	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.113***	-0.120***	-0.133***		
	(0.013)	(0.013)	(0.013)		
Married	0.132***	0.127***	0.105***	0.035	0.027
	(0.016)	(0.016)	(0.016)	(0.039)	(0.038)
Year dummy 2004	0.098***	0.110***	0.111***	0.147***	0.149***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Year dummy 2006	0.201***	0.214***	0.218***	0.290***	0.293***
	(0.016)	(0.016)	(0.015)	(0.023)	(0.023)
Constant	0.881***	1.144***	1.099***	1.584***	1.493***
	(0.047)	(0.049)	(0.047)	(0.136)	(0.133)
Observations	9074	9074	9074	9074	9074
R-squared	0.252	0.274	0.307	0.093	0.109
Number of id				4306	4306

Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
The regressions also include seven regional dummies and eight branch activity dummies.

Table A2. Mean Earnings Regressions for Men

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled OLS	(2) Pooled OLS	(3) Pooled OLS	(4) Fixed Effects	(5) Fixed Effects
Informal Worker		-0.194***		-0.109***	
		(0.021)		(0.033)	
Informal Self-Emp. Worker			0.115***		0.201***
			(0.029)		(0.055)
Informal Wage Worker			-0.163***		-0.066*
			(0.023)		(0.039)
Formal Self-Emp. Worker			0.421***		0.345***
			(0.037)		(0.068)
Years of schooling	-0.020**	-0.019**	-0.036***		
	(0.008)	(0.008)	(0.008)		
Years of schooling squared	0.004***	0.004***	0.005***		
	(0.000)	(0.000)	(0.000)		
Potential experience	0.031***	0.029***	0.028***	0.027**	0.027**
	(0.003)	(0.003)	(0.003)	(0.012)	(0.011)
Potential experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.120***	0.117***	0.099***	0.017	0.009
	(0.024)	(0.023)	(0.023)	(0.043)	(0.042)
Year dummy 2004	0.097***	0.107***	0.108***	0.151***	0.154***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Year dummy 2006	0.202***	0.213***	0.217***	0.297***	0.297***
	(0.020)	(0.020)	(0.019)	(0.029)	(0.029)
Constant	0.880***	1.078***	1.045***	1.606***	1.501***
	(0.064)	(0.065)	(0.063)	(0.179)	(0.173)
Observations	5004	5004	5004	5004	5004
R-squared	0.264	0.278	0.317	0.109	0.130
Number of id				2366	2366

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
The regressions also include seven regional dummies and eight branch activity dummies.

Table A3. Mean Earnings Regressions for Women

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled OLS	(2) Pooled OLS	(3) Pooled OLS	(4) Fixed Effects	(5) Fixed Effects
Informal Worker		-0.306***		-0.197***	
		(0.023)		(0.035)	
Informal Self-Emp. Worker			0.017		0.042
			(0.033)		(0.079)
Informal Wage Worker			-0.330***		-0.178***
			(0.029)		(0.048)
Formal Self-Emp. Worker			0.378***		0.251***
			(0.040)		(0.088)
Years of schooling	-0.005	-0.011	-0.026***		
	(0.010)	(0.010)	(0.010)		
Years of schooling squared	0.003***	0.003***	0.004***		
	(0.001)	(0.001)	(0.001)		
Potential experience	0.025***	0.023***	0.021***	0.028**	0.025**
	(0.002)	(0.002)	(0.002)	(0.013)	(0.013)
Potential experience squared	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.146***	0.141***	0.110***	0.068	0.060
	(0.023)	(0.022)	(0.022)	(0.091)	(0.090)
Year dummy 2004	0.100***	0.112***	0.114***	0.143***	0.145***
	(0.022)	(0.022)	(0.022)	(0.024)	(0.024)
Year dummy 2006	0.198***	0.211***	0.217***	0.282***	0.289***
	(0.025)	(0.024)	(0.024)	(0.040)	(0.040)
Constant	0.749***	1.062***	1.005***	1.563***	1.510***
	(0.071)	(0.074)	(0.070)	(0.251)	(0.248)
Observations	4070	4070	4070	4070	4070
R-squared	0.241	0.273	0.303	0.084	0.092
Number of id				1955	1955

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
The regressions also include seven regional dummies and eight branch activity dummies.

Table A4. Pooled Quantile Earnings Regressions For All Workers

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled .10	(2) Pooled .25	(3) Pooled .50	(4) Pooled .75	(5) Pooled .90	(6) Pooled .10	(7) Pooled .25	(8) Pooled .50	(9) Pooled .75	(10) Pooled .90
Informal Worker	-0.234***	-0.232***	-0.225***	-0.253***	-0.305***					
	(0.024)	(0.020)	(0.020)	(0.018)	(0.033)					
Informal Self-Emp. Worker						-0.134***	-0.069**	0.080***	0.190***	0.296***
						(0.039)	(0.027)	(0.028)	(0.036)	(0.042)
Informal Wage Worker						-0.224***	-0.203***	-0.204***	-0.259***	-0.260***
						(0.029)	(0.021)	(0.022)	(0.024)	(0.037)
Formal Self-Emp. Worker						0.114**	0.298***	0.432***	0.582***	0.697***
						(0.055)	(0.034)	(0.035)	(0.042)	(0.047)
Years of schooling	-0.043***	-0.042***	-0.029***	0.009	0.030**	-0.043***	-0.051***	-0.045***	-0.015**	-0.006
	(0.009)	(0.009)	(0.007)	(0.008)	(0.012)	(0.009)	(0.008)	(0.007)	(0.006)	(0.011)
Years of schooling squared	0.005***	0.005***	0.004***	0.002***	0.001	0.005***	0.006***	0.005***	0.004***	0.003***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Potential experience	0.019***	0.026***	0.027***	0.026***	0.028***	0.020***	0.025***	0.025***	0.024***	0.027***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Potential experience squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.142***	-0.149***	-0.115***	-0.089***	-0.123***	-0.159***	-0.149***	-0.117***	-0.111***	-0.140***
	(0.021)	(0.016)	(0.017)	(0.018)	(0.029)	(0.026)	(0.018)	(0.017)	(0.016)	(0.020)
Married	0.115***	0.101***	0.086***	0.145***	0.178***	0.102***	0.086***	0.068***	0.101***	0.129***
	(0.028)	(0.023)	(0.019)	(0.020)	(0.035)	(0.024)	(0.021)	(0.020)	(0.021)	(0.027)
Year dummy 2004	0.099***	0.077***	0.100***	0.135***	0.156***	0.094***	0.080***	0.118***	0.129***	0.161***
	(0.024)	(0.017)	(0.018)	(0.017)	(0.025)	(0.023)	(0.018)	(0.017)	(0.018)	(0.026)
Year dummy 2006	0.183***	0.181***	0.210***	0.235***	0.263***	0.180***	0.192***	0.226***	0.230***	0.250***
	(0.024)	(0.021)	(0.018)	(0.020)	(0.034)	(0.023)	(0.022)	(0.020)	(0.018)	(0.028)
Constant	0.597***	0.857***	1.077***	1.401***	1.758***	0.594***	0.873***	1.098***	1.291***	1.444***
	(0.078)	(0.059)	(0.058)	(0.062)	(0.115)	(0.074)	(0.056)	(0.056)	(0.057)	(0.099)
Observations	9074	9074	9074	9074	9074	9074	9074	9074	9074	9074

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
The regressions also include seven regional dummies and eight branch activity dummies.

Table A5. Fixed Effects Quantile Earnings Regressions For All Workers

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) FE .10	(2) FE .25	(3) FE .50	(4) FE .75	(5) FE .90	(6) FE .10	(7) FE .25	(8) FE .50	(9) FE .75	(10) FE .90
Informal Worker	-0.195***	-0.181***	-0.152***	-0.121***	-0.103***					
	(0.015)	(0.009)	(0.002)	(0.010)	(0.015)					
Informal Self-Emp. Worker						0.022	0.072***	0.143***	0.221***	0.283***
						(0.020)	(0.015)	(0.005)	(0.015)	(0.020)
Informal Wage Worker						-0.159***	-0.138***	-0.110***	-0.077***	-0.056***
						(0.019)	(0.012)	(0.005)	(0.012)	(0.017)
Formal Self-Emp. Worker						0.193***	0.262***	0.326***	0.395***	0.466***
						(0.022)	(0.021)	(0.005)	(0.021)	(0.027)
Potential experience	0.031***	0.028***	0.027***	0.026***	0.024***	0.031***	0.027***	0.026***	0.025***	0.022***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
Potential experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.028	0.028***	0.035***	0.042***	0.048***	0.018	0.027**	0.027***	0.033***	0.040**
	(0.017)	(0.010)	(0.002)	(0.012)	(0.015)	(0.015)	(0.011)	(0.003)	(0.009)	(0.017)
Year dummy 2004	0.184***	0.167***	0.147***	0.140***	0.131***	0.188***	0.165***	0.149***	0.141***	0.132***
	(0.015)	(0.012)	(0.002)	(0.008)	(0.015)	(0.013)	(0.011)	(0.003)	(0.009)	(0.013)
Year dummy 2006	0.320***	0.298***	0.290***	0.285***	0.292***	0.314***	0.296***	0.293***	0.291***	0.289***
	(0.017)	(0.012)	(0.003)	(0.010)	(0.015)	(0.016)	(0.011)	(0.004)	(0.010)	(0.015)
Constant	1.185***	1.411***	1.584***	1.734***	1.962***	1.127***	1.339***	1.493***	1.646***	1.817***
	(0.030)	(0.025)	(0.004)	(0.026)	(0.030)	(0.035)	(0.024)	(0.006)	(0.026)	(0.032)
Observations	9074	9074	9074	9074	9074	9074	9074	9074	9074	9074

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
The regressions also include seven regional dummies and eight branch activity dummies.

Table A6. Pooled Quantile Earnings Regressions for Men

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled .10	(2) Pooled .25	(3) Pooled .50	(4) Pooled .75	(5) Pooled .90	(6) Pooled .10	(7) Pooled .25	(8) Pooled .50	(9) Pooled .75	(10) Pooled .90
Informal Worker	-0.150***	-0.154***	-0.193***	-0.237***	-0.297***					
	(0.032)	(0.027)	(0.027)	(0.025)	(0.047)					
Informal Self-Emp. Worker						-0.016	0.065	0.126***	0.173***	0.330***
						(0.047)	(0.045)	(0.039)	(0.045)	(0.051)
Informal Wage Worker						-0.148***	-0.103***	-0.145***	-0.228***	-0.230***
						(0.038)	(0.035)	(0.029)	(0.027)	(0.040)
Formal Self-Emp. Worker						0.202**	0.368***	0.424***	0.599***	0.744***
						(0.087)	(0.053)	(0.055)	(0.059)	(0.053)
Constant	0.475***	0.868***	1.039***	1.357***	1.624***	0.438***	0.828***	1.065***	1.301***	1.348***
	(0.090)	(0.081)	(0.076)	(0.084)	(0.153)	(0.085)	(0.091)	(0.071)	(0.072)	(0.125)
Observations	5004	5004	5004	5004	5004	5004	5004	5004	5004	5004

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.The regressions also include the set of control variables present in Tables A1 to A5 plus seven regional dummies and eight branch activity dummies.

Table A7. Pooled Quantile Earnings Regressions for Women

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) Pooled .10	(2) Pooled .25	(3) Pooled .50	(4) Pooled .75	(5) Pooled .90	(6) Pooled .10	(7) Pooled .25	(8) Pooled .50	(9) Pooled .75	(10) Pooled .90
Informal Worker	-0.324***	-0.342***	-0.295***	-0.266***	-0.277***					
	(0.038)	(0.032)	(0.029)	(0.034)	(0.052)					
Informal Self-Emp. Worker						-0.275***	-0.175***	0.063	0.209***	0.364***
						(0.065)	(0.048)	(0.048)	(0.049)	(0.057)
Informal Wage Worker						-0.388***	-0.359***	-0.322***	-0.320***	-0.227***
						(0.063)	(0.046)	(0.043)	(0.044)	(0.054)
Formal Self-Emp. Worker						0.020	0.235***	0.415***	0.585***	0.736***
						(0.078)	(0.067)	(0.053)	(0.055)	(0.059)
Constant	0.431***	0.663***	0.960***	1.428***	1.714***	0.480***	0.678***	1.045***	1.293***	1.346***
	(0.132)	(0.073)	(0.119)	(0.128)	(0.141)	(0.125)	(0.085)	(0.092)	(0.112)	(0.108)
Observations	4070	4070	4070	4070	4070	4070	4070	4070	4070	4070

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.The regressions also include the set of control variables present in Tables A1 to A5 plus seven regional dummies and eight branch activity dummies.

Table A8. Fixed Effects Quantile Earnings Regressions for Men

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) FE .10	(2) FE .25	(3) FE .50	(4) FE .75	(5) FE .90	(6) FE .10	(7) FE .25	(8) FE .50	(9) FE .75	(10) FE .90
Informal Worker	-0.132***	-0.129***	-0.109***	-0.087***	-0.073***					
	(0.018)	(0.012)	(0.004)	(0.014)	(0.021)					
Informal Self-Emp. Worker						0.125***	0.138***	0.201***	0.265***	0.295***
						(0.027)	(0.018)	(0.008)	(0.021)	(0.025)
Informal Wage Worker						-0.095***	-0.085***	-0.066***	-0.051***	-0.043*
						(0.020)	(0.013)	(0.008)	(0.014)	(0.025)
Formal Self-Emp. Worker						0.246***	0.280***	0.345***	0.415***	0.463***
						(0.030)	(0.028)	(0.009)	(0.029)	(0.033)
Constant	1.154***	1.444***	1.606***	1.766***	2.000***	1.084***	1.346***	1.501***	1.665***	1.879***
	(0.044)	(0.030)	(0.010)	(0.032)	(0.042)	(0.050)	(0.032)	(0.014)	(0.029)	(0.043)
Observations	5004	5004	5004	5004	5004	5004	5004	5004	5004	5004

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.The regressions also include the set of control variables present in Tables A1 to A5 plus seven regional dummies and eight branch activity dummies.

Table A9. Fixed Effects Quantile Earnings Regressions for Women

Dependent Variable: Log Hourly Real Earnings

VARIABLES	(1) FE .10	(2) FE .25	(3) FE .50	(4) FE .75	(5) FE .90	(6) FE .10	(7) FE .25	(8) FE .50	(9) FE .75	(10) FE .90
Informal Worker	-0.254***	-0.238***	-0.197***	-0.165***	-0.128***					
	(0.024)	(0.015)	(0.004)	(0.017)	(0.022)					
Informal Self-Emp. Worker						-0.103***	-0.034	0.045***	0.124***	0.225***
						(0.030)	(0.027)	(0.009)	(0.026)	(0.034)
Informal Wage Worker						-0.259***	-0.232***	-0.176***	-0.126***	-0.089***
						(0.029)	(0.021)	(0.010)	(0.023)	(0.028)
Formal Self-Emp. Worker						0.107***	0.190***	0.253***	0.324***	0.442***
						(0.040)	(0.036)	(0.009)	(0.033)	(0.044)
Constant	1.211***	1.403***	1.563***	1.718***	1.897***	1.166***	1.361***	1.508***	1.645***	1.791***
	(0.047)	(0.035)	(0.008)	(0.042)	(0.049)	(0.045)	(0.038)	(0.012)	(0.042)	(0.041)
Observations	4070	4070	4070	4070	4070	4070	4070	4070	4070	4070

Bootstrapped standard errors (150 replications) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.The regressions also include the set of control variables present in Tables A1 to A5 plus seven regional dummies and eight branch activity dummies.