EARNING STRUCTURE AND HETEROGENEITY OF THE LABOR MARKET: EVIDENCE FROM DR CONGO

Douglas Amuli Ibale

LIDAM DISCUSSION PAPER | IRES/2020/37

UCLouvain

IRES | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

Earning structure and heterogeneity of the labor market: Evidence from DR Congo*

Douglas AMULI Ibale[a]

[a] Institut de Recherche Économique et Sociale (IRES-UCLouvain (Belgium))

[a] Laboratoire d'Economie Appliqué au Développement (LEAD-UCB/ DR Congo)

December, 2020

Abstract

Using a unique broad individual, household and expenditure survey data on the DRC, initial descriptive statistics highlight five different sectors on the labor market with two "higher-paid" that are completely formal and two "lower-paid" that are largely informal. Based on a linear regression result, we report a significant heterogeneity across them when it comes to earnings. With an unconditional quantile regression methodology corrected for selectivity bias we show that, though the effect of education on earnings provides a clear support to the human capital theory, basic education has no significant impact on earnings in higher-paid sectors. Likewise, tertiary education matters for earnings in lower-paid sectors as well. We then decompose the earning gap across sectors and show that workers of the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. However, when higher-paid and lower paid sectors are concerned, the coefficient effect at the upper end of the distribution is negative. Implying that the labor market provides an "informal employment earning premium" to some workers of the lower-paid sectors whose, given their characteristics, wouldn't do better in the higher-paid sectors.

Keywords: Earning, Labor market, Heterogeneity. Earning decomposition.

JEL codes: E26, J21, J31, J4 , J7, J82

*We would like to thank Frédéric Docquier, Muriel Dejemeppe and two anonymous referees for their invaluable comments. We also thank Celestin Bucekuderhwa Bashige and the participants at doctoral workshops at UCLouvain and Universite Catholique de Bukavu for helpful suggestions. We are grateful towards UCLouvain for supporting this research through the Fond de cooperation au development". The usual disclaimers apply. Correspondence: Douglas AMULI Ibale (douglas.amuliibale@uclouvain.be)

1 Introduction

Due to sectoral wage determination process, labor market heterogeneity in developing economies has received a great deal of attention in the literature. This stems from the fact that, it exhibits considerable diversity that includes substantial segments (hereafter called sectors) of different characteristics (Fields, 2009) that depend on the specific environment in which workers operate (Deakin, 2013). Though in a general neo-classical framework, the level of earning and earnings of the workforce is determined by demand for and supply of labor, in developing countries, this is affected largely by strict labor market sectors and strong entry barriers across different sectors of the labor market. In self-employment activities, workers enjoy non-earning features such as greater flexibility (in terms of working hours, work relationships, responsibilities, etc.) and maximize more their utility then their earnings. In the public sector, earnings are mainly determined through a political process or service regulations (Gunderson, 1979). In the private formal sector, earnings are determined by the demand and supply conditions of the labor market and in the private informal sector, earnings are said to be low and volatiles (Günther and Launov, 2017). These differences bring a natural interest in the type and the size of the labor market's sectors and mainly, their earning outcomes. Because these outcomes are a robust indicator of the livelihood status of the workforce and hence, of the populations who derive all or the great bulk of their earnings from their work on the labor market.

Drawing on an empirical case study on DRC, this paper provides insights on the functioning of labor markets in Sub-Saharan African (SSA) countries. With a selectivity-corrected Mincerian earning equation, we investigate the role of individuals characteristics in determining their earnings. An Oaxaca-Blinder decomposition of the earning gap between sectors is carried out, allowing us to disentangle differences in characteristics (e.g. education) and difference in the return on these characteristics across sectors as drivers of the earning gap. Our initials descriptive statistics give five different sectors with two higher-paid completely formal sectors (Public and Private formal) and two lower-paid largely informal sectors (Self-employment and Private informal) that are significantly heterogeneous across them when it comes to earnings. Based on earning equations, we show that, though the effect of education on earnings provides a clear support to the human capital theory, basic education has no significant impact on earnings in higher-paid sectors. Likewise, tertiary education matters for earnings in lower-paid sectors as well. We then decompose the earning gap across sectors taken in pairs and show that workers of the lower-paid sectors earn less not only because they are less skill-endowed but also because they earn lower returns on such skills. However, when higher-paid and lower-paid sectors are concerned, the coefficient effect in the upper end of the distribution is negative, implying that the labor market provides an *informal employment earning premium* to some individuals of the lower-paid sectors whose, given their characteristics, would not do better in the higher-paid sector.

We conduct our analysis on the Democratic Republic of Congo (DRC), a typical example of an economy where labor market is badly out of balance. Galloping demography is continuously increasing the demand for jobs while, since 1990 firm's labor demand has been falling steeply as a result of looting, wars and other shocks to the economy. This state of affairs combined with high poverty, high unemployment[1] and inexistent public provisions for unemployment insurance has favored the attractiveness of public jobs and the emergence of the informal sector where self-employment dominates. As a consequence, in the formal sector, salaries are negotiated in a context of strong demand for employment, 78.4% of formal firms report to compete with informal ones (World-Bank, 2005) and the wage bill constitutes now the largest item in public sector spending as the scantiness of private jobs and the economic and political instability have fostered the rush for public jobs. Accordingly, the country is gaining insights into the factors influencing the self/paid-employment selection, the public/private sector selection, and sectoral earning determination processes for paid employees. This provides rich evidence of the labor market heterogeneity that stands out as an interesting case for analysis in the SSA context.

Due to important consequences on economic growth, poverty and inequality, the economic literature has intensively explored the heterogeneity of the labor market and earning differences between type of workers. For example, labor market and earnings display substantial heterogeneity with respect to gender (Oaxaca, 1973; Polachek and Xiang, 2009), education levels (Beaudry and David, 2003), self/paid-employment (Bernhardt, 1994) and to public/private sector (Christofides and Panos, 2020; Tansel, 2005). There is also a positive formal-informal earning differential (Falco et al., 2010) which is explained both by individuals characteristics and by the return to these characteristics (Garcıa, 2017). The limit of these studies is that, they concentrate on a dual analysis where market is a matter of two to three sectors, which is a quite narrow selection of employment opportunities in the context of SSA. Furthermore, they completely exclude from the analysis a substantial number of individuals that are outside the labor market and can be defined as unemployed.

In this paper, we provide the first study on labor market heterogeneity and earning differential in DRC. The contribution of this study is twofold: First, we begin with the structure of the labor market and then link earnings to that structure. In so doing, we let workers decide first whether to enter the labor force by accounting for the unemployment

[1]See IMF (2015)'s country report for detailed statistics on unemployment and poverty in DRC

segment. Indeed, given the high rates of unemployment in developing countries, on the one hand it points to the fact that being at the bottom end of the earnings distribution may not be the worse outcome for those in the labor market. On the other hand, many individuals either work in the household for no explicit pay, or as unpaid apprentices or are simply paid in kind. Therefore workers choose between paid and unpaid work in first stage, rather than between sectors as in the aforementioned studies. Second, we allow individuals to have a larger variety of choices and combinations on the labor market. For instance, they choose whether to work as self-employed or as wage employees and within wage employment whether in the public or the private sectors. Within the private sector, they can also choose between the formal or the informal sector. Accounting for all the possible sectors is important because high degree of heterogeneity imposes greater resource costs on the economies, especially these of developing countries, by causing the failure of the market to move the 'right' resources into 'right' sectors (Berry and Sabot, 1978). Of course, as some of these sectors choices are constrained by rationing, workers may not necessarily end up where they wish. We then examine the earning differential and its decomposition across the observed sectors taken in pairs for a detailed investigation. With such a comprehensive analysis, we aim to help developing countries to understand how do labor market operate, to identify the extent of its heterogeneity and how the latter is responsible for labor market outcomes and distortions that lead to earning differentiation across workers.

In the next sections, we outline the data used and give some descriptive statistics. In section 3 we explain our estimation procedure before presenting and discussing our results in section 4. Section 5 concludes.

2 Data and descriptive statistics.

In 2005 and 2012, the DRC's National Institute of Statistics (INS) conducted, in partnership with different actors (including Afristat and the World Bank), a broad household and expenditure survey that followed the so-called 1.2.3 methodology. Each of these numbers refers to a collection phase. Phase 1 provides detailed information on employment, unemployment, household and individual socio demographic characteristics. Phase 2 specifically gathers information on the characteristics of firms and firm owners in the informal sector. Phase 3 is a survey on household expenditures. In this paper, our sample focuses on the phase_1 of the 2012 survey limited to individuals between the age 15 and 65 years involved or not on the labor market. The sample covers about 88 600 individuals from 11 Provinces of the DRC. The survey provides data on five sectors of the labor market. 1) Individuals out of the labor market; this sector includes both the participants and non-participants. The participants are unemployed who are looking for employment and are available upon request. The non-participants are unemployed individuals unavailable for work, due to studies, disabilities, early retirement and others. 2) The self-employed[2], who at the time of the interview, work in a business that they was owned entirely or partly. 3) Public workers hired by the public administration. 4) Private formal employees and 5) Private informal employees. Such data gives an opportunity to study the heterogeneity of the labor market and the resulting earning differential. To check whether these sectors are different when it comes to earnings, we run a simple linear regression with monthly log-earning as the dependent variable. The independent variables are the different labor market sectors namely; self-employment, public workers, private formal and private informal workers entered as dummies. In this paper, we use monthly earnings instead of hourly earnings[3]. Given the often constrained working hours in the informal and self-employment sectors, monthly earnings reflect better earning opportunities in some sectors than hourly earnings.

In Table 1, the summary statistics provided reveal that there are considerable differences among workers employed in these sectors. It shows that self-employment and unemployment sectors are markedly large. Indeed, as previously said, the unemployment sectors includes non-participants and participants. This last category represents 8.5% of the total unemployed individuals. For the self employment sector, its magnitude comes from the fact that it includes all individuals with small business and street or market vendors who work for themselves. More than 90% of this sector is made up of informal entrepreneurs. As regard to earnings, that is the monthly earning of the main activity held in the last 30 days, the data show a remarkable difference in average earnings across the different sectors. Indeed, Table 1 shows that the national average monthly earning is CDF 51 946.8 . Furthermore, before the change that occurred in 2018 , the minimum wage in DRC amounted to CDF 1680 per day (i.e., a monthly earning of CDF 50 400). If we consider the minimum wage used by the "Institut National de la Statistique" (INS, 2014), after updating the minimum wage by taking into account the annual increase implemented by the law and the increase in the general price level, the minimum wage in DRC is set to 54 128 CDF. Taking the two statistics, we see that there is, on the one side two "higher-paid" sectors, namely public and private formal sectors, where the average monthly earning is above the national average and far above the updated minimum wage. Individuals in the two higher-paid sectors are 100% formal workers. On the other side, we have two "lower-paid" sectors, namely Self-employment and Private informal sectors,

[2]Includes independents with and without wage employee.

[3]Indeed, many activities in the informal as in self employment sector are part-time jobs by nature and employees in those sectors could not easily increase their earning by simply providing more labor hours per month.

where the average monthly earning is below the national average and even bellow the updated national minimum wage. Furthermore, individuals in the two lower-paid sectors are more than 95% informal workers.

This variation in earning between higher and lower-paid sectors is explained by some important differences between sectors. For example, in regard to education, we see that public and private formal workers are more endowed for secondary and tertiary education than workers in others sectors of the labor market. Indeed, on average, higher-paid workers have two times the years of schooling compared to lower-paid-sectors. Private informal workers are the less educated even compared to the unemployed. Low skilled individuals escape unemployment in informal sector, while most high skilled ones often prefer to stay unemployed until they find a formal job. This is stated by Reyes et al. (2017), skilled individuals could have the ability to sustain a longer job search to find a more suitable match. Indeed, those with higher education typically have the economic backing from their families to engage in a longer job search. Moreover, unemployed, self-employed and private informal workers are more likely to be female and less educated, while most self-employed and private informal workers live in rural areas. For other variables, the Table 1 shows that, individuals whose father works as public worker or as self-employed are more likely to be public servant or self employed. The descriptive statistics show that 48.63% of self-employed individuals have a father working as self-employed. We do not test the differences between these characteristics across the different sectors of the labor market. This is because if earning is different across sectors, the cross earning gap can result from how the labor market discriminates between workers of same characteristics across different sectors or how the labor market reward individuals of different characteristics across sectors.

3 Estimation strategy.

We want to examine the heterogeneity in the labor market and study the earning differential between workers across sectors of the labor market. To start with, we estimate the following equation in order to illustrate earnings differential across sectors i[4] of the labor market.

$$Y_i = \lambda_o + \lambda_1 SE_i + \lambda_2 PS_i + \lambda_3 PFS_i + \lambda_4 PIS_i + \epsilon_i \quad (1)$$

Y_i stand for the monthly earning in log form, SE; Self-Employment, PS; Public sector, PFS; Private Formal Sector, PIS; Private Informal Sector and ϵ is the error term.

Next, we examine earning differential across sectors and explore the determinants of

[4]Here, we only consider four sectors by excluding the unemployment sector as it not concerned by earnings.

monthly earnings within each sector by regressing monthly earnings on education, controlling for personal and geographical characteristics. We rely on a quantile decomposition methodology (as introduced by Koenker and Bassett (1978)), allowing us to analyze the earning gap across sector at different points of the distribution. We follow Firpo et al. (2009)'s method to compute unconditional quantile and decompose the effect of each covariate at different parts of the distribution. Unconditional Quantile Regression (UQR) uses the influence function[5] as the dependent variable at different quantiles. In the case of quantile q_τ of an outcome variable Y and explanatory variables X, the The Recentered Influence Function[6] (RIF) is

$$RIF(Y; Q_\tau) = Q_\tau + \frac{\tau - I(Y \leqslant Q_\tau)}{f_Y(Q_\tau)} \tag{2}$$

Where $f(.)$ is the density and $1(.)$ indicates that Y is at or above the quantile q_τ. It is recentered as q_τ is added and as a consequence, the expected value of the RIF is q_τ itself. Indeed, Firpo et al. (2009) show that this property extends to the conditional-on-controls RIF and the earning equation of the RIF model at quantile τ, with $\tau \in (0, 1)$, is then:

$$RIF(Y; Q_\tau) = X\beta_\tau + \nu_\tau \tag{3}$$

Where Y is the natural logarithm of monthly earning and X is a vector of K explanatory variables (including the constant), β_τ is the corresponding coefficient vector and ν_τ is the corresponding error term. As explanatory variables in the earning equation we include the variables usually included in a Mincerian earning regression-education (These are dummy variables for the highest completed education of the individual; basic, secondary and tertiary education) and control for gender, age (and age square), experience, and dummies of job trained and leaving in urban areas. Furthermore, dummy variables for province of residence are included to control for differentials in cost of living and labor-market opportunities.

[5]The influence function measures how robust a distributional statistic is to outliers. it is given by : $IF(Y; Q_\tau) = \tau - I(Y \leqslant Q_\tau)/f_Y(Q_\tau)$, where $I(Y \leqslant Q_\tau)$ is an indicator function taking value one if the condition in (.) is true, zero otherwise.

[6]Since the explanatory variables do not enter into the transformation of equation (2), although the X's in the model change, the interpretation of the estimated effects does not vary, and so alternative models can be compared and different sources of socioeconomic inequality incorporated. The main advantage of this method over conditional regression is that the estimated effects do not depend on the set of explanatory variables in the model. Moreover, as in the conditional regression, the estimates are robust to outliers.

3.1 Accounting for selection

The distribution of workers among the four different sectors is not random. In estimating the earning equations, the selection into different sectors for which we observe earnings must be taken into account. Potential biases could result from ignoring sample selection (Heckman, 1974). To take this into account, we assume that individuals face five mutually exclusive choices: Unemployed ($j = 0$), Self-employment ($j = 1$), public administration employee ($j = 2$), Private formal employee ($j = 3$) and Private informal employee ($j = 4$). Worker's tastes and preferences as well as human capital and other characteristics will determine the sectoral choice. We assume a conditional multinomial logit model for the probability that the individual chooses alternative j relative to that of being in an arbitrarily chosen reference sector (which is non-employment) as follows:

$$P_j = exp(Z\alpha_j)/\left[1 + \sum_{j=1}^{4} exp(Z\alpha_j)\right] \tag{4}$$

Where Z is is a vector of explanatory variables affecting sectoral choice and α_j is a vector of unknown parameters of the alternative j. The selection equation is in reduced form, in the sense that the earning rate is not included as an explanatory variable and we only consider someone's actual state. Information on preferred labor market state or job search is not taken into account. We adopt the two-step estimation method developed by Lee (1983) and Trost and Lee (1984). In the first stage, we estimate the sectoral choice probabilities by maximum likelihood logit method and construct the selection term for the alternative j as follows:

$$\lambda_j = \phi(H_j)/\Phi(H_j) \tag{5}$$

Where $H_j = \Phi^{-1}(P_j)$, ϕ is the standard normal density function, and Φ is the standard normal distribution function. In the second stage, the estimated λ_j is included among the explanatory variables of the earning equations. The implied earning equations are then estimated using a RIF regression, providing consistent estimates of the parameters. The term λ_j plays the same role as Mill's ratio in the usual Heckman (1979) procedure but it is here quantile-specific Töpfer (2017). A statistically significant positive value of the self-selection term indicates that persons who – after controlling for observable characteristics – are more likely to work in sector j also have, ceteris paribus, higher expected wages in this sector (see Dimova et al. (2011)). A negative estimate, by contrast, would indicate that persons who are more likely to work in sector j have lower expected wages in this sector.

In order to achieve identification of the parameters in the multinomial logit, we introduce in its equation variables that influence labor-force participation and sector choice but may be excluded from the earning equations. We include; unearned income (whether a family member owns shares, securities or financial investments), risk aversion (proxied by whether the individual has a health insurance or not), whether or not a close member of family lost his job and dummies of father working in the public sector and father working as self-employed. These variables provide labor-market information and should be associated with labor market participation but not earnings. For instance, as suggested by Schultz (1990), unearned income is expected to reduce the probability of participation by raising the shadow value of a person's time in non-market activities and in self-employment. A risk averse individual will prefer being employed to working as self-employed or being unemployed (Christofides and Panos, 2020). Having a father working in the public sector may increase the probability of employment of an individual in the public administration sector, and similarly for the other variables and sectors. For these variables, the relevance of the exclusion restrictions in terms of their predictive power of sector choice can be directly tested from the model estimates. However, no formal over-identification test has been developed for this specific framework. We are aware of the fact that, as usual, the validity of our exclusion restrictions is debatable, because it can be argued that the selected variables might be related to unobserved determinants of earnings. This would be especially true in the case that the list of control variables in the earning equation(s) does not include all the relevant features of the current sector. Nevertheless, in Table 6 in the appendixes we provide an informal means of testing both for the relevance of the exclusion restrictions and for the excludability of the aforementioned variables from the outcome equations.

3.2 Decomposition strategy

The earning gap between workers of different sectors can exist not only because of disadvantages in terms of remunerated characteristics, but also because the returns to these characteristics can be different for workers of different sectors. To assess the earning gap between sectors, the Oaxaca (1973) and Blinder (1973)'s decomposition methodology is implemented. Based on separate estimation of earning equations, the Oaxaca and Blinder's decomposition generates the counterfactual on the basis of which the difference in average earnings between sectors is broken into two additive components: one attributable to differences in average characteristics of the individuals (the characteristics effect), and the other to the differences in the rewards associated with these characteristics (the coefficient effect). The method by Firpo et al. (2009) allows to conduct a detailed Oaxaca and Blinder's type decompositions using unconditional quantile regression that accounts for selection (Töpfer, 2017). Indeed, the earning gap is, as in the standard two-fold Oaxaca-Blinder decomposition, decomposed into characteristics (explained) and coefficients (unexplained) components. The decomposition for the τth quantile $\hat{\triangle}_\tau = \overline{RIF}_j$ - $\overline{RIF}_i$ takes the form hereunder:

$$\underbrace{\hat{\triangle}_\tau}_{\textit{Earning gap}} = \underbrace{\bar{X}_j - (\bar{X}_i)\hat{\beta}_j}_{\textit{Characteristics effect}} + \underbrace{\bar{X}_i(\hat{\beta}_j - \hat{\beta}_i)}_{\textit{Coefficient effect}} \quad (6)$$

Where $\overline{RIF}_j$ and $\overline{RIF}_i$ are the estimated mean $\widehat{RIF}$ of sectors j and i respectively. Taking one sector (the sector with high mean income) as the reference category, the method has the advantage that it computes detailed decomposition and allows for the unconditional mean interpretation of the coefficient estimates. When the earning gap is mainly attributable to the coefficient effect, earning differences are attributable to differences in returns on skills between sectors. When, the earning gap is primarily explained by the characteristics effect, earning differences between sectors are due to differences in workers' endowments. In our estimation, we use a bootstrap procedure with fifty replications to estimate standard errors for the estimated coefficients, for earning gap as well as for the characteristics and coefficient effects.

4 Estimation Results.

In this section, we present the results obtained from the Multinomial logit to examine the probability of being in a given sector, the regression result of earning's determinant in each sector of the labor market on the $10th$, $50th$ and $90th$ percentiles and the results of the earning decomposition across the different sectors of the labor market. But first, the results of equation (1) where y stand for the monthly earning in log form, and each sector is entered as dummy variable are presented in Table 2. The last four columns of Table 2 represent different reference groups for the four sectors of the labor market where workers get monthly earnings. Given the significance of the coefficients, the regression results show that the four sectors are statically different from each other, when monthly earning is considered. A separate earning function is then needed for each of the sectors to capture the sector-specific earning determinants.

4.1 Multinomial Logit Estimates

Multinomial logit estimates of sector choice are shown in Table 3. The table gives the marginal effects of each variable on the probability of joining a particular sector calculated at the mean values of the variables, with the unemployment sector as the reference group. The results indicate that, males are less likely to work in private informal sector but more likely to be in the three others sectors, particularly in the self-employment. Basic education does not matter for working in the formal private sector meanwhile it decreases the probability of being employed in the private informal and the public sectors compared to unemployed individuals. However, individuals with basic education are more likely to work as self-employed. On the one hand, secondary and tertiary education significantly increase the probability of working in the private formal and in the public sector. Furthermore, the higher the education level the higher its contribution to the participation in the two aforesaid sectors. On the other side, these individuals with secondary and tertiary education are less likely to work informally or being self-employed and the higher the education level the higher its probability of no participation in the private informal or self-employed sector. As the two sectors, namely self-employment and informal employment are the lower-paid sector of the labor market, educated individuals tend to avoid them while queuing for public and private formal jobs. Except for being employed in the private formal sector, experience and marital status significantly increase the probability of employment in all the others three sectors and mainly the self-employment sector. Probably that, with more experience, private formal workers who have accumulated wealth prefer to shift and work as self-employed.

The job training significantly increases the probability of employment in all of the four sectors, with more weight for the self-employment sector.Urban individuals are more likely to be private formal workers than unemployed. This can be explained by the fact that formal firms are more located in urban areas. Income effects on participation are measured by the unearned income of the individuals. The result indicate that individuals with unearned income are more likely to be private formal or public workers than unemployed. As hypothesized previously, risk aversion individuals are less likely to be self-employed. This said, the risk aversion significantly decreases by 8.7 percentage point the probability of working as self-employed but, it increases significantly the probability to participate on the labor market as wage employee. If a family member lost his job, individuals are more likely to work informal. This is more likely if the lost family member was providing income to the household, the remaining unemployed members of the family should find faster how to compensate for the revenue lost. The easiest way is then to go informal as the entry barriers are weak in this sector. As expected, having a father working as public servant significantly increases the probability of participation as employee in the public administration or working as self-employed. This effect is however, negative for participation in the private informal and nil in the private formal sector. Similarly, individuals with parents working as self-employed are more likely to work as independent, but less likely to work in the private sector. In the second step, we used the estimates from the multinomial logit regression as a selection equation to address the selection bias into the wage equations.

4.2 The Wage Equations

Selectivity-corrected estimates of the sectoral wage equations for employed of different sectors are given in Tables 4 and 5, respectively. All of the wage equations are statistically significant overall. From these tables, the results for the unconditional quantile regressions at the sector level show that mostly the selection terms "δ" are statistically significant in all sectors. These results indicate a presence of sample selection bias for individuals across the earning distribution in the sectors of the labor market. Tables 4 and 5 summarize the results for quantile regressions at the 10th, 50th and 90th percentiles for each sector. Focusing on the gender variable, the results highlight a significant earning gap between men and women across all four sectors of the labor market that have earnings and the gap appears to be larger in lower-paid sectors. For example, in self-employment sector, man's expected earnings at the bottom of the earning distribution is about 25.2% higher than a woman's expected earnings. This difference goes to 43.9% at the top of the earning distribution. In the private informal sector, in the middle quantiles of the distribution, a woman's expected earnings at the 50th percentile are approximately 58.8% lower than a man's expected earnings and 48.6% lower at the 90th percentile.

Sidelined by the scarcity of positions in higher-paid sectors given their characteristics, many males have joined the lower-paid sectors and have become more competitive in sectors often reserved to their female counterparts. Of course, similar discrimination is found for higher-paid sectors but with lower intensity. In general, whatever the sector, the results show more pronounced barriers for women competing for jobs. Linear and quadratic terms in age have the expected positive and negative signs respectively, in all sectors. This implies that age dividend does exist in all the four sectors with more weight in the middle of the distribution in the private informal sector. In the latter tier of the distribution, one additional year increases for 10,2% the monthly earnings of individuals. In regard to education, basic and secondary education do matter for earnings only in lower-paid sectors. Indeed, as previously said, the self-employment sector is made up of individuals with small business and street vendors where more than 90% of them are informal self-employed individuals. The two sectors, self-employment and private informal sectors are thus of lower productivity where workers are said to face low and volatile earnings in small firms with labor-intensive activities and without job security. These facts make the two sectors weakly attractive to tertiary educated individuals but more attractive to low educated individuals who barely have room in higher-paid sectors. However, the tertiary education significantly increases earnings in all the sectors. In higher-paid sectors, tertiary education increases by 68.1 percentage point the monthly earning in the public sector while it almost double (97.6 percentage point) the monthly earning of individuals in the middle tier of the private formal sector. Overall, tertiary educated people have advantage over the rest whatever the sector of employment. The higher earnings associated with age, education provides clear support to the human capital theory in the public and private sector (Becker, 1964; Mincer, 1974). However, the relative low/insignificant returns to basic education in higher-paid sectors that are known to be legal as they operate formally erodes the individual's real incentive to obtain low levels of education and questions the quality of schooling in DRC. Furthermore, higher levels of education tend to yield significant positive returns in lower-paid sectors (largely informal and less productive) as well. For a country with a relatively high level of informality, this is area calls for further attention and investigation. It is known that informality decreases with education. But, on the one side, if there is no earning premium to basic education in higher-paid (that are formal sectors) labor sectors in a country where it is so costly to study, people may choose working underground instead. On the other side, if education, both basic and tertiary levels are rewarded in the lower-paid (mostly informal sectors) as well the latter, as more flexible, with less entry and exit barriers and without or less tax burden, will be attracting more and more educated people eroding the country's productivity and wealth in the long run. The ending point could be a vicious circle of poverty where the country and its population are caught in an informality trap.

Surprisingly job training effect, thought positive, is insignificant in the private formal sector. But this can probably be explained by the presence of highly educated individuals in this sector. Table 1 showed that 38.53% of workers of that sector have tertiary education. This level of education might act as job training and as seen previously, higher education influence strongly the workers' earnings. The significant returns to job training in the public sector can be liked with benefit in the form of incremental salary, additional allowance, or promotion in accordance with the government policy. Concerning individual experience, its positive effect is significant only in higher-paid sectors. Finally, living in urban areas often procures higher returns to workers regardless of the sector of employment on the labor market.

4.3 Decomposition Results

For a better-detailed analysis and a clear and simplified view, this section presents the decomposition results in the form of graphs[7]. Figure 1 plots the estimated wage gap, correcting for selection for six cases representing the interaction of four sectors of the labor market taken in pairs. Taking the sector with high mean income as the reference sector in each pair, the six pairs of sectors are as follow; a) Earnings in public sector vs in self-employment, b) Earnings in private formal sector vs in self-employment, c) Earnings

[7]Estimates can be made available upon request.

in self-employment vs in private informal sector, d) Earnings in private formal sector vs in public sector, e) Earnings in public sector vs in private informal sector and f) Earnings in private informal sector vs in private informal sector. Equal weights are assigned to the two sectors for each pair. Figure 1 contains six subfigures and reports the earning gap of the six pairs.

Considering the subfigures A, B, E and F where a higher-paid sectors is the reference sector over a lower-paid sector, we observe that the estimated earning gap is positive throughout the earning distribution. This confirms that workers in higher-paid sectors are more remunerated than their counterpart in the lower-paid sectors. However, the earning gap is not homogeneous throughout the earning distribution. In the bottom part of the distribution it is large (this is more pronounced when the public sector is the reference sector). Around the fifth percentile, it goes decreasing to be lower in the second half of the distribution. This heterogeneity in the distribution of the earning gap indicates that not all individuals in the lower-paid sectors are the same compare to individuals in the higher-paid sectors when it comes to earnings. Focusing on each set of factors, the coefficient effect dominates the characteristic effect in the bottom of the distribution. Subsequently, much of the earning gap at the bottom of the distribution result from workers in the lower-paid sectors being remunerated less than workers of the higher-paid sectors for the remunerated characteristics. The coefficient effects decrease over the distribution, whereas the characteristic effects increase in the second half of the distribution, particularly toward the upper end of the distribution where characteristics effects dominate the coefficient effects. This fact indicates that on the one side, workers of the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. On the other side, individuals at the top end of the distribution in lower-paid sectors earn less compare to their counterpart of the higher-paid sectors because the latter have superior skills.

This position of workers in the lower-paid sectors vis-a-vis of workers in the higher-paid sectors shows two different types of lower-paid workers in the DRC's labor market. First, those at the bottom of the distribution, who, despite *identical* (or almost) characteristics to their counterparts' workers at the bottom of the distribution in the higher-paid sectors, earn less on such characteristics. They represent the disadvantaged group, working in the "easy-entry" segment of the lower-paid sectors. Second, workers of lower-paid sectors at the upper tiers of the distribution, where the characteristics effect explains much of the earning gap. In this upper tier, we observe a relative lower earning gap between workers of the higher-paid and lower-paid sectors. Both types of workers could have no significant differences in the rates of return on characteristics but the difference in earnings is due to differences in skills. This group of workers may be associated with the advantaged segment of lower-paid sectors and might be linked with entrepreneurship. Largely informal, these workers endowed with some particular skills prefer the combination of monetary rewards and greater flexibility (in terms of working hours, work relationships, responsibilities, etc.) in the informal sector (Fields, 1990). In this advantaged tier of the distribution, lower-paid individuals might be enjoying non-earning features and maximize more their utility rather than their earnings. More interestingly, in the top end of the distribution, the coefficient effect is negative in the four subfigures. This informs that, in this tier of the distribution, moving individuals of the lower-paid sector to higher-paid sectors and rewarding them according to their characteristics could even lower their actual earnings. We then have a "*lower-paid sector employment earning premium*" on the labor market. Due to the prevalence of informal individuals in the lower paid sector as previously showed, the lower-paid sector employment earning premium is indeed an "*informal employment earning premium*" for individuals who, given their characteristics, wouldn't do better in the higher-paid (formal) sector. In this context, the significant rationing of higher-paid jobs and relative abundance of informal workers, particularly those with very low qualification levels, undermine the benefits of higher-paid sectors for low-skilled individuals.

In the Subfigure D of Figure 1 , the pair of sector concerns the two higher-paid sectors of the labor market where private formal is the reference sector. The figure shows a positive earning throughout the earning distribution implying that Private formal workers earn more than Public workers. Both for this pair and the previous ones, the earning gap is not homogeneous along the distribution. However, differently from the previous case, the earning gap is lower at the bottom of the distribution and higher in the large part of the second half of the distribution. As public jobs are financed by means other than those operating in the private sector and that, earnings are mainly determined through a political process or service regulations (Gunderson, 1979), earning disparities between lower and higher earners within the public sector are less important. However, in the private formal sector, productivity matters for earnings. This fact generally leads to a large earning disparity between lower and higher productive workers (earners) within the private formal sector. This difference in the determinants of earnings between the two higher-paid sectors leads to an increasing earning gap between them throughout the distribution. As regard to each set of factors, the coefficient effect exceeds the characteristics effect for 80% of the earning distribution. From this, we learn that the earning difference between the two higher-paid sectors in DRC is mainly due to the difference in skill remuneration rather than skill composition with a higher return in the Private formal sector. Finally, the pair of lower-paid segment is given in the Subfigure D of Figure 1. Due to the relative lower earning gap, we conducted, in addition to the result of the linear regression in Table

2 that showed a significant difference between the two lower-paid sectors, a mean test difference of the earnings between them and a likelihood-ratio test for combining alternatives sectors in Table 7. Overall, the result showed that their mean earnings are significantly different and that the two lower-paid sectors are significantly distinguishable with respect to the variables in the model. Thus, no categories should be combined. Given that, we did separate the two sectors in this decomposition section. Between the two lower-paid sectors, a relative earning gap, mainly explained by the skill remuneration, exists in the first half of the distribution but disappears in the second half. This positive gap can be due to the presence of formal entrepreneurs in the Self-employment sectors whose skill remuneration in the formal sector makes the difference. However, as one moves up along the earning distribution, the earnings in Self-employment is caught by that of Private informal sector. The vanishing earning gap can be partially explained by the fact that formal entrepreneurs who are the higher earners of the self-employment sector are paying tax on their higher earnings meanwhile higher earners of the Private informal sector are not. In any cases, the vanishing gap is consistent with greater freedom of choice between Self-employment and Informal job as individuals move up along the distribution. With a marginal earning gap in this tier of the distribution, workers in the Private informal sector may to some extent be willing to accept lower earnings to avoid the administrative cost of social security in regulated sectors, when it is perceived as costly and ineffective Garcıa (2017).

5 Conclusion

In this paper, we take advantage of a large-scale survey in DRC to examine the heterogeneity in the labor market and examine the earning gap and its decomposition across different sectors throughout the earning distribution. Using a Mincerian selectivity-corrected sectoral wage equations estimated for each sector, the role of workers' personal characteristics is explored in the first step. In the second step, an Oaxaca-Blinder decompositions of the wage differentials between sectors are carried out. Initial data description reports five different sectors in the DRC labor market. A basic earning regression with four sectors entered as dummies shows clearly that they are statistically different when earnings are concerned. The analysis of sector earnings combined with the stylized facts of DRC's labor market allows to group the four sectors into two mains group. On one side two higher-paid sectors, fully made up of formal individuals, where the average monthly earning is above the national average and far above the updated national minimum wage. On the other side two lower-paid sectors, mainly made up of informal individuals, where the average monthly earning is below the national average and far below the updated national minimum wage.

The earning functions show individuals self-select in different sectors of the labor market. The higher earnings associated with age, education and others provides a clear support to the human capital theory. The result shows that basic education is not significantly rewarded in higher-paid sectors, meanwhile tertiary education matter for earnings in lower-paid sectors as well. The decomposition results report that, when higher-paid sectors are taken as reference sectors, the earning gap is positive but not homogeneous throughout the earning distribution. Furthermore, this positive earning gap is due both to skill remuneration in the bottom part of the distribution and skill composition in the upper tier of the distribution implying that, workers of the lower-paid sectors earn less not only because they are less skill endowed but also because they earn lower returns on such skills. More interestingly, when the pair of sector concerns higher-paid vs lower-paid sectors, the coefficient effect is negative in the upper tier of the distribution, highlighting an informal employment earning premium on the labor market as lower-paid sectors are mainly informal. When the two higher-paid sectors are concerned, private formal workers earn more than public workers. However, differently from the previous case, the earning gap, mainly attributable to skill remuneration, is lower at the bottom of the distribution, but higher in the large part of the second half of the distribution. When the two lower-paid are concerned, the earning gap appears only in the first half of the distribution and vanishes in the second half, undermining the benefits of Self-employment.

Our results have policy implications for development strategies, in DRC and the developing world that aim, whether to combat earning discrimination on the labor market for well-being purpose, whether to safeguard diversity of sectors in the labor market with specific regulations, for entrepreneurial spirit and household subsistence strategies purposes. The coexistence of diverse sectors on the labor market with differences in earning outcomes calls for various strategies to face the large share of lower-paid sector and to limit informality that, in some circumstances, provide a premium to workers but increases discrimination on the labor market. Specific policies have to be constructed for each particular group or sector on the labor market. This is important because, on the one side we show that different mechanisms may be working at the group of sector level or even at each sector level and on the other side. because most people, especially the poor, derive all or the great bulk of their income from the work they do.

References

Beaudry, P. and David, G. (2003). Wages and employment in the united states and germany: What explains the differences? *American Economic Review*, 93:573–603.

Becker, G. (1964). *Human Capital*. Columbia University Press: New York.

Bernhardt, I. (1994). Comparative advantage in self-employment and paid work. *Canadian Journal of Economics*, 27:273–289.

Berry, A. and Sabot, R. (1978). Labour market performance in developing countries: a survey. *World Development*, 6(12):1199–1242.

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Christofides, Louis, N. and Panos, P. (2020). Self/paid-employment. public/private sector selection and wage differentials. *Labor Economics*, 9(6):737–762.

Deakin, S. (2013). Addressing labour market segmentation: The role of labour law' governance. *Centre for Business Research: Working Paper*, 52.

Dimova, R., Gang, I., and Landon-Lane, J. (2011). Revealed informal activity. *IZA Discussion Papers*, 5607.

Falco, P., Kerr, A., Rankin, N., Sandefur, J., and Teal, F. (2010). The returns to formality and informality in urban africa. *CSAE Working Paper 3, Oxford*.

Fields, G. (1990). Labour market modelling and the urban informal sector: Theory and evidence. In *In D. Thurnham, B. Salome, and A. Schwarz : The Informal Sector Revisited, (eds), Paris: OECD*. OECD.

Fields, G. (2009). Segmented labor market models in developing countries. In *The Oxford Handbook of Philosophy of Economics*. Oxford University Press, New York.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77:953–973.

Garcıa, G. A. (2017). Labor informality: Choice or sign of segmentation? a quantile regression approach at the regional level for colombia. *Review of Development Economics*, 21(4):985–1017.

Green, C., Heywoody, J., and Theodoropoulos, N. (2014). Performance pay and ethnic earnings differences in britain. *Oxford Economic Papers*, 66(3):798–823.

Gunderson, M. (1979). Earnings differentials between the public and private sectors. *Canadian Journal of Economics*, 12:228–242.

Günther, I. and Launov, A. (2017). Informal employment in developing countries. *Journal of Development Economics*, 97:88–98.

Heckman, J. J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, 42:679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

IMF (2015). Democratic republic of the congo, country report. Technical Report 15/281, International Monetary Fund.

Koenker, R. and Bassett, Jr., G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(3):545–61.

Mincer, J. (1958). Investments in human capital and personal income distribution. *Journal of Political Economy*, 56:281–302.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. NBER and Columbian University Press: New York.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Polachek, S. and Xiang, J. (2009). The gender pay gap across countries: A human capital approach. *SOEP papers on Multidisciplinary Panel Data Research*, 227.

Reyes, A., Alvaro, G., Dino, M., Carly, P., and Javier, S.-R. (2017). Democratic republic of congo: Jobs diagnostic. Technical report, World Bank CC BY 3.0 IGO.

Schultz, T. (1990). Testing the neoclassical model of family labor supply and fertility. *Journal of Human Resources*, 25:599–634.

Tansel, A. (2005). Public private employment choice. wage differentials. and gender in turkey. *Economic Development and Cultural Change*, 53(2):453–477.

Trost, Robert, P. and Lee, L.-F. (1984). Technical training and earnings: A polytomous model with selectivity. *Review of Economics and Statistics*, 66(2):151–56.

Töpfer, M. (2017). Detailed RIF decomposition with selection: The gender pay gap in italy. *Working Paper*, 26.

World-Bank (2005). Enterprise survey. Technical report, World Bank. http://www.enterprisesurveys.org/Data/ExploreTopics/informality.

Appendix

Table 1: Descriptive statistics

Variables	Total	Unemployed	Self-employed	Public workers	Private Formal workers	Private Informal workers
Size	55.652	21043	22331	3289	1069	7920
Monthly earning (CDF)	51946.8	-	42954.26	94901.43	178185.8	41451.6
Male (Proportion)	48.40	44.94	51.96	78.47	78.92	32.50
Age(year)	32.82	25.89	37.70	41.95	38.25	32.58
Education(year)	6.76	7.68	5.62	11.87	12.41	4.92
Education (Proportion)						
-Illiterate	21.66	14.05	27.94	5.02	2.89	34.63
-Primary school	24.61	21.31	30.06	4.72	4.01	29.96
-Secondary school	47.59	57.47	40.22	63.52	54.57	33.03
-Tertiary school	6.14	7.17	1.78	26.73	38.53	2.38
Job training (Proportion)	0.05	0.00	3.17	51.26	41.14	5.67
Experience (year)	8.18	1.28	13.35	11.83	7.99	10.43
Living in urban (Proportion)	50.54	66.89	35.97	66.52	93.17	35.80
Marital (Proportion)	58.98	29.57	78.59	82.82	72.22	70.11
Head of hh (Proportion)	34.85	9.82	55.21	77.96	72.40	20.97
Others(proportion)						
Unearned income (Proportion)	1.45	1.90	0.79	3.05	6.06	0.93
Father in the public sector(Proportion)	18.84	21.93	14.61	35.94	36.95	13.02
Father is Self-employed (Proportion)	40.42	33.28	48.63	26.79	16.09	45.16
Risk averse (Proportion)	4.79	6.90	1.91	11.21	18.38	3.13
Family member lost a job (Proportion)	2.55	3.18	1.93	2.09	3.54	2.73

Notes: Authors computation using 1.2.3 data

Table 2: Linear regression

Variables	(1)	(2)	(3)	(4)
Constant	10.06***	10.97***	11.63***	9.97***
	(0.007)	(0.018)	(0.032)	(0.012)
Self employment	-	-.909***	-1.57***	.088***
	-	(0.019)	(0.033)	(0.014)
Public sector	.909***	-	-.667***	.997***
	(0.019)	-	(0.037)	(0.022)
Private-Formal sector	1.57***	.667***	-	1.665***
	(0.033)	(0.037)	-	(0.034)
Private-Informal sector	-.088***	-.997***	-1.665***	-
	(0.014)	(0.022)	(0.034)	-
Observation	33241			
$F(3, 33237)$	1459.92			
Prob $\succ$ F	0.0000			
Adj R-squared	0.1164			

Notes: Authors computation. Notes: Log monthly earnings is the dependent variable. ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis. In regression (1), (2), (3) and (4), Self-employment, Public, Private Formal, and Private Informal sectors are respectively taken as the base sector.

Table 3: Determinants of sectors' choice

	Self-employed		Public workers		Private formal workers		Private informal workers	
	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
Male	0.0753***	(0.0036)	0.0010***	(0.0017)	0.0081***	(0.0013)	-0.0825***	(0.0030)
Age	0.0232***	(0.0009)	0.0010***	(0.0004)	0.0012***	(0.0003)	-0.0054***	(0.0007)
Age_sq	-0.0003***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)
Education								
Primary	0.0224***	(0.0051)	-0.0100***	(0.0027)	0.0014	(0.0016)	-0.0184***	(0.0043)
Secondary	-0.0157***	(0.0049)	0.0286***	(0.0024)	0.0130***	(0.0014)	-0.0557***	(0.0042)
Tertiary	-0.1324***	(0.0098)	0.0653***	(0.0034)	0.0287***	(0.0024)	-0.0656***	(0.0087)
Experience	0.0135***	(0.0002)	0.0004***	(0.0001)	-0.0003***	(0.0001)	0.0058***	(0.0002)
Job training	0.3137***	(0.0645)	0.0343***	(0.0014)	0.0140***	(0.0015)	0.2114***	(0.0277)
Living in urban	-0.0258***	(0.0039)	-0.0281***	(0.0019)	0.0212***	(0.0021)	-0.0563***	(0.0034)
Marital	0.0741***	(0.0040)	0.0082***	(0.0019)	-0.0006	(0.0013)	0.0150***	(0.0034)
Unearned income	0.0187	(0.0177)	0.0035***	(0.0047)	0.0054**	(0.0026)	-0.0244	(0.0160)
Risk aversion	-0.0874***	(0.0104)	0.0102***	(0.0026)	0.0079***	(0.0015)	0.0177**	(0.0084)
Member of family lost his job	-0.0466***	(0.0111)	-0.0140***	(0.0045)	-0.0025	(0.0029)	0.0330***	(0.0087)
Father is a public servant	0.0227***	(0.0051)	0.0099***	(0.0017)	-0.0008	(0.0012)	-0.0258***	(0.0045)
Father is an Self-employed	0.0534***	(0.0038)	0.0004***	(0.0019)	-0.0052***	(0.0015)	-0.0073**	(0.0031)
Province fix effect	Yes		Yes		Yes		Yes	
Number of observation	57 204							
Prob $\succ$ F	62692.57***							

Notes: Notes: Author's calculations based on 1.2.3 data. The reference sector is the unemployment sector. ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis. Illiterate is the excluded categories in education. All models contain Province dummies.

Table 4: Quantile regressions : Lower-paid sectors

	Self-employment			Private Informal workers		
	10	50	90	10	50	90
Male	0.252***	0.410***	0.439***	0.227***	0.588***	0.486***
	(0.038)	(0.023)	(0.027)	(0.047)	(0.053)	(0.056)
Age	0.071***	0.065***	0.054***	0.073***	0.102***	0.067***
	(0.010)	(0.006)	(0.007)	(0.010)	(0.010)	(0.009)
Age_sq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Primary	0.102**	0.091***	-0.002	0.075	0.214***	0.118***
	(0.046)	(0.027)	(0.026)	(0.052)	(0.054)	(0.039)
Secondary	0.288***	0.303***	0.217***	0.083	0.461***	0.470***
	(0.045)	(0.027)	(0.029)	(0.057)	(0.059)	(0.053)
Tertiary	0.547***	0.587***	1.210***	0.194**	0.976***	1.423***
	(0.092)	(0.071)	(0.144)	(0.076)	(0.109)	(0.224)
Experience	-0.001	0.001	0.001	-0.005*	-0.012***	-0.004
	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Job training	0.226***	0.238***	0.354***	-0.064	0.274***	0.566***
	(0.075)	(0.055)	(0.089)	(0.062)	(0.074)	(0.126)
Formal	-0.222***	-0.348***	-1.186***	-	-	-
	(0.055)	(0.126)	(0.333)	-	-	-
Without Employee	0.003	-0.075***	-0.104***	-	-	-
	(0.047)	(0.028)	(0.034)	-	-	-
Urban	0.188***	0.270***	0.425***	0.185***	0.458***	0.485***
	(0.037)	(0.024)	(0.029)	(0.047)	(0.054)	(0.052)
δ	0.052***	0.015**	-0.008	0.021***	0.059***	0.048***
	(0.011)	(0.006)	(0.009)	(0.008)	(0.010)	(0.012)
Constant	8.317***	9.421***	12.713***	7.589***	8.196***	10.607***
	(0.268)	(0.291)	(0.697)	(0.188)	(0.190)	(0.206)
Province fe	Yes	Yes	Yes	Yes	Yes	Yes
Prob $\succ$ F	33.51***	91.65***	64.61***	19.16***	115.04***	41.03***
Number of obs		21732			7216	

Note: ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis

Table 5: Quantile regressions : Higher-paid sectors

	Public workers			Private Formal workers		
	10	50	90	10	50	90
Male	0.032	0.032***	0.218***	0.077	0.163**	0.135*
	(0.073)	(0.013)	(0.063)	(0.122)	(0.082)	(0.074)
Age	0.021	0.014***	0.040***	0.064*	0.059***	0.024
	(0.022)	(0.003)	(0.015)	(0.037)	(0.020)	(0.019)
Age_sq	0.000	0.000***	0.000**	-0.001	-0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Primary	-0.483**	-0.156***	-0.144	0.827	-0.252	-0.050
	(0.240)	(0.035)	(0.127)	(0.520)	(0.232)	(0.131)
Secondary	-0.126	-0.061**	0.040	1.001**	-0.328*	0.006
	(0.163)	(0.026)	(0.106)	(0.441)	(0.189)	(0.098)
Tertiary	-0.069	0.057**	0.681***	0.997**	0.103	0.422***
	(0.177)	(0.029)	(0.135)	(0.451)	(0.202)	(0.120)
Experience	0.022***	0.003***	0.005	-0.010	0.006*	0.007**
	(0.004)	(0.001)	(0.004)	(0.008)	(0.004)	(0.004)
Job training	-0.154**	0.036***	0.179***	0.012	-0.009	0.024
	(0.078)	(0.012)	(0.062)	(0.103)	(0.071)	(0.085)
Urban	0.976***	0.098***	0.423***	1.410***	0.476***	-0.059
	(0.087)	(0.013)	(0.045)	(0.352)	(0.113)	(0.077
δ	0.050**	-0.001***	0.015**	0.014	0.028***	0.028***
	(0.020)	(0.002)	(0.007)	(0.014)	(0.010)	(0.007)
Constant	9.359***	10.648***	10.963***	7.064***	10.281***	12.030***
	(0.569)	(0.083)	(0.348)	(0.958)	(0.526)	(0.436)
Province fe	Yes	Yes	Yes	Yes	Yes	Yes
Prob $\succ$ F	14.82***	61.03***	20.56***	6.08***	36.34***	6.81***
Number of obs		3248			1043	

Note: ***,**,* Denote significance at 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis

Figure 1: Earning differentials between formal and informal sector over different quantiles of the earning distribution by province

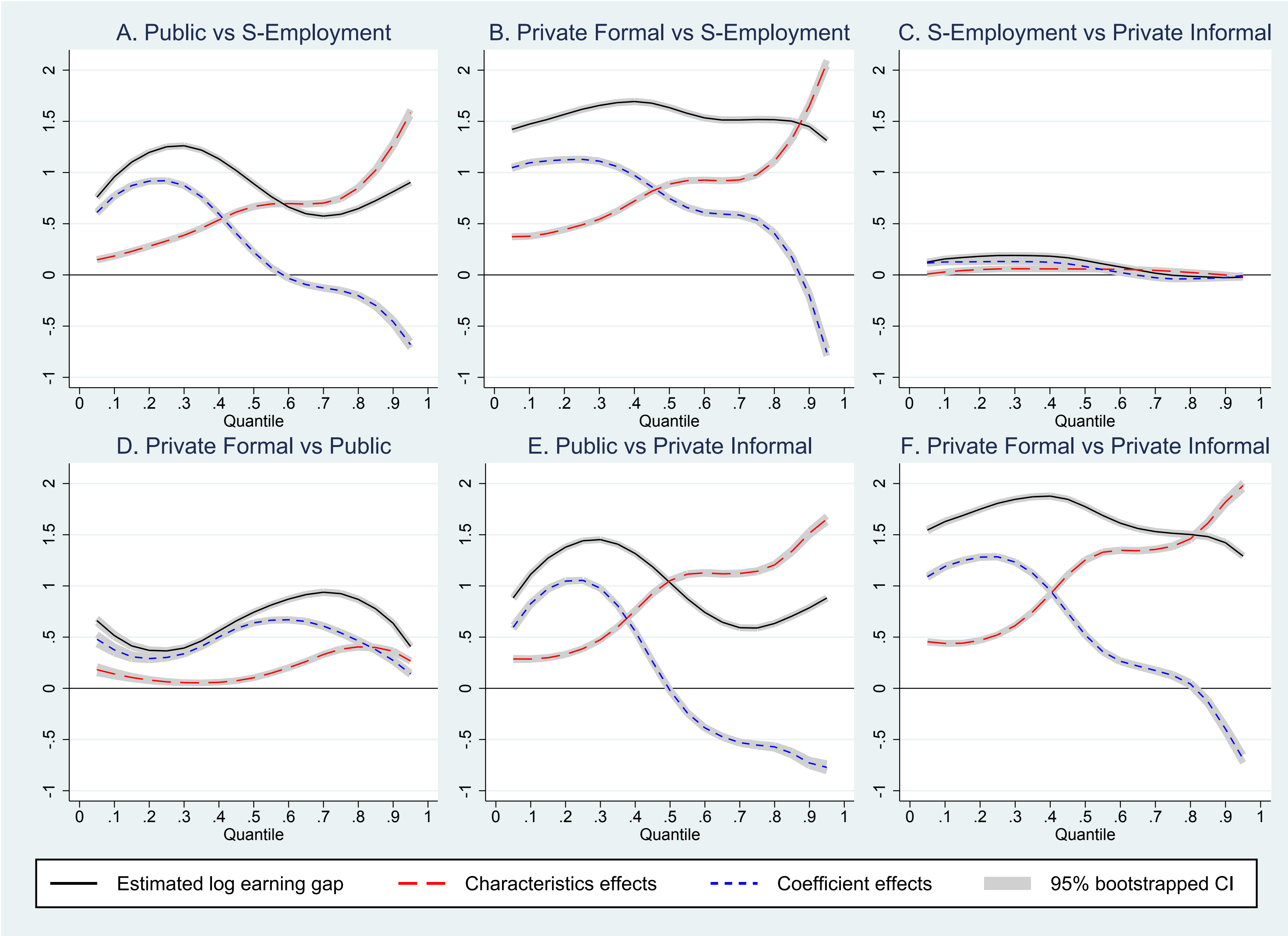

Table 6: Testing the validity of the exclusion restrictions

Sectors	Relevance	Excludability		
		10	50	90
	Wald Test ($P-value$)	Wald Test ($P-value$)	Wald Test ($P-value$)	Wald Test ($P-value$)
Unemployed	364.15 (0.0000)	– –	– –	– –
Self-Employed	364.15 (0.0000)	2.64 (0.0216)	0.90 (0.4812)	1.00 (0.4180)
Public	73.89 (0.0000)	1.22 (0.2979)	1.60 (0.1584)	0.20 (0.9632)
Private_Formal	23.23 (0.0003)	0.63 (0.6420)	0.79 (0.5595)	0.40 (0.8046)
Private_Informal	121.66 (0.0000)	0.98 (0.4320)	2.07 (0.0674)	0.27 (0.9283)

Note: Relevance: *statistical significance of the exclusion restrictions in the five sector choice equations.*
Excludability: *statistical insignificance of the exclusion restrictions in each earning equation.*

Here, we seek to provide evidence of the validity of the elicited exclusion restrictions of variables incorporated in the multinomial model for identification. For them to be valid, the variables should be relevant determinants of sector choices but not directly related to earnings, once we have conditioned for the employment sector and other attributes. The validity of the "relevance" condition can be directly tested from the estimates of the multinomial selection equation. Table 6 above contains several Wald tests for the joint statistical significance of the exclusion restrictions for each estimated model. As can be seen, the relevance of the exclusion restrictions for the whole multinomial model is clearly not rejected by the data. Taking each equation separately, the variables included are good predictors of the differences in the likelihood of each sector (in relation to be unemployed), As for the "excludability" condition, no formal overidentification test has yet been developed in this framework. Therefore, this condition has to be informally checked by examining the joint statistical significance of the exclusion restrictions in the outcome equation(s) of sectors, conditional on other determinants of earnings. The results of these Wald tests — which are reported in the four last column of Table 6 - suggest that the exclusion restrictions are not jointly significant in the outcome equations at any conventional significance level, with the exception of Self-employed sector on the tenth percentile and private informal sector on the fifteenth percentile, in which the null hypothesis that the exclusion restriction's coefficients are jointly equal to zero is not rejected when considering a significance level of 5%. Overall, the evidence obtained when adopting this informal approach to demonstrating the validity of the exclusion restrictions suggests that the model is well identified.

Table 7: Test of significance

Mean difference test

Groups	Observation	Mean	P-value
Self-Employment	22331	10.061	
Private informal	7920	9.97	0.0000
Difference		0.088	

Based on the above, There is a significant mean difference of log earnings between the two sectors

Likelihood-Ratio test for combining alternatives

	chi2	df	P≻ chi2
Unemployed & Self-Employment	36224.949	26	0.000
Unemployed & Public	27430.368	26	0.000
Unemployed & Private Formal	9773.842	26	0.000
Unemployed & Private Informal	22323.603	26	0.000
Self-Employment & Public	13218.431	26	0.000
Self-Employment & Private Formal	4590.221	26	0.000
Self-Employment & Private Informal	3720.946	26	0.000
Public & Private Formal	1323.106	26	0.000
Public & Private Informal	10228.398	26	0.000
Private Formal & Private Informal	3675.629	26	0.000

Based on the above, No categories should be combined

INSTITUT DE RECHERCHE
ÉCONOMIQUES ET SOCIALES

Place Montesquieu 3

1348 Louvain-la-Neuve

ISSN 1379-244X D/2020/3082/37

IRES | Louvain Institute of Data Analysis and Modeling in economics and statistics