

# **Double Penalty in Returns to Education: Informality and Mismatch in the Labor market, Evidence From Colombia. \***

**Paula Herrera-Idárraga \*\***

AQR Research Group – University of Barcelona and Pontificia Universidad Javeriana.

**Enrique López-Bazo**

AQR Research Group – University of Barcelona and European Commission, Joint Research Center (JRC), Institute for Prospective Technological Studies (IPTS).

**Elisabet Motellón**

Universitat Oberta de Catalunya and AQR Research Group – University of Barcelona

**Preliminary version February 2013**

## **Abstract**

The aim of this paper is to examine the returns to education taking into consideration the existence of educational mismatches in the formal and informal employment of a developing country. For this purpose we estimate the standard Duncan and Hoffman's specification (so called ORU wage equation) at the mean and at different quantiles, using quantile regression estimation, and controlling for a rich set of observable individual and firm characteristics exploiting a micro-data from Colombia. In both cases we correct for the endogeneity of sector participation. Our results show that the returns of surplus, required and deficit years of schooling are different in the two sectors. Moreover, these returns vary along the wage distribution and the pattern of variation is not the same for formal and informal workers. In particular, we find that informal workers not only face lower returns to their education, but that there is a second penalty associated with educational mismatches that puts informal workers at a greater disadvantage compared to their formal counterparts.

*JEL classification:* O17; J21; J24

*Keywords:* Educational Mismatch; Formal/Informal Employment; Wage Gap; Economic Development

\***Acknowledgements:** The authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation, *Plan Nacional de I+D+i*, Project ECO2011-30260-C03-03, and the European Commission under the 7<sup>th</sup> Framework Programme project SEARCH, SSH-2010-2.2-266834. Paula Herrera-Idárraga also acknowledges financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (grant FI-DGR 2012) and from the Pontificia Universidad Javeriana.

\*\***Corresponding author:** Paula Herrera-Idárraga. Mailing address: Diagonal, 690, 08034 Barcelona, Spain; telephone number: +0034-934031907; e-mail: paulaherrera@ub.edu.

## **1. Introduction**

A distinctive feature of almost all Latin American and the Caribbean labor markets is the existence and the persistence of a large informal sector. In fact, half of the employed population of this region worked in informal jobs at the end of the first decade of this century (ILO's 2011 Labour Overview). Informal employment embraces a variety of heterogeneous activities, such as self-employment entrepreneurs, salaried workers of large and small firms, and unpaid domestic workers. Informal employment generally involves that workers are trapped in unproductive activities, with inferior working conditions, lack of social security and lower earnings. A seemingly stylized fact, found in past studies about labor market segmentation, is that informal-sector workers, even if equally productive, are subject to lower remuneration than formal-sector workers. So even when more highly educated workers tend to be more productive than less skilled counterparts, education may not be the key for higher paying jobs if the labor market is segmented.

A number of explanations have been offered to explain why some earning-relevant characteristics, for example, education, are better rewarded in the formal sector than in the informal sector. An important bulk of these explanations is based on a segmented view of the labor market. For instance, the presence of extremely restrictive labor market institutions and strict regulation of entry into the formal sector could pose a possible cause, so that some workers that do not have access to the formal sector are forced to accept informal sector jobs characterized by inferior earnings (see Fields, 1975). However, several more recent studies postulate, for both firms and workers the decision of being formal turns out to be extremely costly, due to the non labor costs associated with health and pension contributions, payroll taxes, commuting subsidies, among others, which significantly increases the attractiveness of informal activities. Maloney (1999), for instance, introduces a standpoint in which workers may find informal-sector employment a desirable alternative, due to inefficiencies in the provision of public services, i.e. health and pension, or because their level of human capital do not fulfill the

requirements for performing a formal jobs. In the last case, a wage penalty for informal-sector employment may be due to sorting, where those with low levels of human capital are also those more likely to work in the informal sector (Tokman, 1982). This type of sorting may result from the fact that firms in the informal sector have limited access to financing and employers choose to substitute low-skill labor for physical capital (see, e.g., Amaral and Quintin 2006).

However, none of the former studies have considered one aspect which can affect the wage gap between formal and informal workers, that is, the way workers match their acquire education to the one required to perform their job. One important feature that raises concern in developed and developing countries is the existence of a discrepancy between the education attainment of workers and the skill requirements of jobs, commonly known as education-occupation mismatch. The incidence and labor market effects of educational-occupation mismatch, especially over-education, have received increasing attention in the literature for developed countries (Germany, the Netherlands, Spain, the United Kingdom, Portugal, Hong Kong and the United States)<sup>1</sup> and recently some attention has also been paid for some developing countries (Mexico, Pakistan, India, the Philippines, Thailand and Colombia)<sup>2</sup>. Several of these studies have shown that the incidence of education-occupation mismatch varies significantly with the method used to measure required education, hence over-education. However while the choice of the method can have an effect on the incidence of the phenomenon under analysis, the effect on earnings is not altered (Groot and Maasen van den Brink 2000). So, independently of the method used, a number of studies that estimated the effects of over-education on earnings for developed and developing countries found that, overeducated workers tend to earn higher returns to their years of schooling than co-workers who are not over-educated, but lower returns

---

<sup>1</sup> Duncan and Hoffman (1981), Verdugo and Verdugo (1989), Sicherman (1991), Tsang, Rumberger and Levin (1991), McGoldrick and Robst (1996) studied the phenomenon for the United States; Alpin, Shackleton and Walsh (1998), Green, McIntosh and Vignoles (2002), Dolton and Vignoles (2000) and Chevalier (2003) for the UK; Hartog and Oosterbeek (1998) and Groot and Massen van den Brink (2000) for Holland; Bauer (2002) and Buchel and van Ham (2003) for Germany; Kiker, Santos and De Oliveira (1997) and Mendes de Oliveira, Santos and Kiker (2000) for Portugal; Alba-Ramirez (1992) for Spain. For an extensive review of overeducation in developed countries see McGuinness (2006) and for a recent survey on overeducation see Leuven and Oosterbeek (2011).

<sup>2</sup> Quinn and Rubb (2006) study the phenomenon for Mexico, Abbas (2008) for Pakistan and Mehta et al. (2011) for India, Mexico, the Philippines and Thailand, Mora (2005), Castillo (2007) and Herrera-Idárraga et al. for Colombia.

than workers with similar education who work in jobs that require the level of education that they possess.

In a previous study, Herrera-Idárraga et al. (2012), using micro-data for Colombia, find that after controlling for other characteristics and correcting for endogeneity, informal salary workers are more likely to be over-educated than formal workers. Thus it is possible that the formal/informal wage gap is driven, at least in part, by a less satisfactory matching of education-occupation in the informal sector and by the penalization in terms of wages that is derived from this mismatch. Actually the aim of this paper is to reexamine the wage gap between formal and informal workers taking into consideration that education-occupation mismatch is present in both sectors, using the case study of Colombia. Colombian labor market constitutes a good case of study for several reasons. First, informality today is at center of economic and political debates in the country because the high levels that prevail. Second, in Colombia there is a high incidence of the minimum wage, i.e., a relatively high proportion of formal sector employees, 34.6%, receive a salary similar to the minimum (Arango, Herrera and Posada 2008) which points toward to the existence of important labor market rigidities. Third, previous studies have found overeducation to exist in Colombia (Mora 2005, Castillo 2007, Dominguez-Moreno 2009 and Herrera-Idárraga et al. 2012).

In particular this study addressed the following questions: Are the return to years of required education, years of surplus education and years of deficit education different across formal and informal sectors? And, to what extent these differences in the returns help to explain the wage gap between formal and informal workers? The idea of distinguishing the difference in the returns from correct, over and under education for the two sectors is a novel contribution, as there is no previous contribution that considered this difference before in all studies of which we know about informality<sup>3</sup>.

---

<sup>3</sup> See, e.g., Magnac (1991), Nuñez (2002), Maloney and Nuñez (2004), Floréz (2002), Kugler and Kugler (2009) and Mondragón-Vélez, Peña and Wills (2010) for Colombia; Gindling (1991) for Costa Rica; Pradhan and van Soest for

The empirical analysis consists of examining the returns to education taking into consideration the existence of educational mismatches in the formal and informal sector. For this purpose we first estimate the standard Duncan and Hoffman's specification (so called ORU wage equation) at the mean, using ordinary least square (OLS), and controlling for a rich set of observable individual and firm characteristics. Then, we examine if the returns to education for each of the education-occupation mismatch are not uniform along the wage distribution by using quantile regression estimation. In both cases the endogeneity sector choice is addressed.

Preliminary results for Colombia show that: i) consistent with previous literature, the return to an overeducated year is lower than the return to a required year of education, both in the formal and informal sector, ii) formal workers that possess the education required to do their job have a higher return to their education, around double, compared with their informal counterparts, iii) moreover, they have a higher return than informal workers who are overeducated, iv) the return to an overeducated year of education is higher in the formal sector than in the informal sector and v) the wage penalty of deficit schooling is almost the same across the two sectors. Moreover using quantile regression estimations we show that i) these returns vary along the wage distribution and ii) the pattern of variation along the distribution is not the same for formal and informal workers. More specifically, the returns to required education increases along the wage distribution for both type of workers, but the increase is more noticeable for formal workers. While to surplus education increases along the wage distribution for formal workers they almost remain constant for informal workers. We therefore conclude that adding measures of educational mismatch gives important information to the analysis of the formal/informal wage gap. In particular, we show that in the informal sector not only the returns to education are lower, but the penalty that informal workers face due to educational mismatches in terms of wages are considerable higher than their formal counterparts.

---

Bolivia (1995); Amuedo-Dorantes (2004) for Chile; Pratap and Quintin for Argentina (2006); Tansel (2000) for Turkey; Marcouiller, Ruiz de Castilla and Woodruff (1997) and Gong and Van Soest (2002) for Mexico; Botelho and Ponczek (2011) for Brazil; Badaoui, Strobl and Walsh (2008) for South Africa.

The rest of the paper is organized as follows. The next section gives a description of the data and some selected descriptive, while the empirical approach is presented in section 3. Section 4 summarizes the results regarding the estimates of the empirical models, and, finally, section 5 concludes.

## **2. Dataset and descriptive analysis**

We use data from the Colombian Household Survey (CHS), a repeated cross-section conducted by the National Statistics Department (DANE), for 2010. The survey gathers information about employment conditions for population aged 12 or more including income, occupation and industry sector at two digit level, in addition to the general population characteristics such as sex, age, marital status and educational attainment and covers the thirteen mayor metropolitan areas in Colombia.

A sample of 34626 of working individuals was drawn from the 2010 CHS. The analysis was restricted to salary workers that were not carrying formal studies aged between 15 and 60 years and who report working more than 16 hours per week. We do not include self-employed and employers workers in the analysis because their source of income is a combination of labor and physical capital and therefore may not be compared with earnings of other employees. Apart from this, self-employed workers' earnings would be expected to have a greater measurement error. Also, while comparing self-employed informal workers to their formal sector counterparts may be of interest in its own right, it has been show in previews studies that self-employed in the informal sector corresponds more with a voluntary entry, while informal salaried work may correspond more closely to the standard queuing view, especially for younger workers (Bosh and Moloney 2010). Excluding self-employed resulted in dropping 16941 individuals. We also exclude form the sample public employees since by nature they belong to the formal sector and their wages might reflect institutional arrangements. After excluding observations with missing

values or inconsistencies for the selected regressors, over 13797 individuals remained in the working sample.

We classify workers as formal or informal according to whether they are covered by the social security system or not, in line with the definition proposed by the Seventeenth International Conferences of Labour Statisticians (ICLS).<sup>4</sup> Thus, we define workers as formal if they contribute both to health and old-age insurance. For the purpose of measuring the incidence of the education-occupation mismatch we define required education using the statistical method in its mean and mode version. Under the statistical method required education is defined as the mean or mode level of schooling for each occupation. Individuals are classified as over-educated (under-educated) for a particular occupation if their level of education is higher (lower) than the required education. In the mean measure a worker is over-educated or under-educated if their completed level of schooling deviates by one standard deviation from the mean in their occupation.<sup>5</sup> Regarding earnings, we have combined information from monthly income and worked hours in order to obtain hourly wages. The basis of our earnings comparison is wages before taxes, which could overestimate the premium associated with the formal sector employment.

Table 1 contains mean hourly wages by job type and educational mismatch. As it can be seen informal workers are likely to earn less than formal workers, formal workers earn almost twice what informal workers earn. One should note that these large earnings differentials found here are in line with several other studies of other countries. The classification of workers by educational mismatch gives an interesting point of view of this wage gap. For instance the wage gap is not uniform across the different types of educational mismatch. While overeducated

---

<sup>4</sup> The definition of the Seventeenth International Conferences of Labour Statisticians (ICLS) of informal employment is "based on the characteristics of the individual's employment, job or position. A worker has an informal job if the employment relationship is, in law or in practice, not subject to national labour or social legislation. This condition of informal employment is observed in persons employed in both formal and informal enterprises, as well as in those employed in domestic service by households". (ILO's 2011 Labour Overview).

<sup>5</sup> For purpose of brevity we only included the results obtained with the mean, as with the mode the results are not significantly different. Results are available on request.

formal workers earn more than twice than their informal counterparts, this wage gap is reduced significantly for correct and undereducated formal workers. This simple descriptive analysis gives some insight of the importance to take into account educational mismatch for understanding the formal/informal wage gap. Table 2 presents some basic summary statistics concerning the distribution of characteristics of our total sample and classified into those working in the formal and in the informal sector that may be driving this earnings differentials. Formal workers in our sample are more likely to have higher education or more (44%), whereas informal workers are more likely to have basic secondary and secondary (22% and 36% respectively). There is not significant difference in the age and experience display in both groups. In contrast, there are some notable differences in the average tenure between sectors; formal workers tend to accumulate much more tenure than informal workers, suggesting high stability of employment. In fact, 95% of formal workers have a signed a contract, and 65% of them of a permanent type, in contrast with only 18% of informal workers who have a contract, and only 10% having a permanent one. On the other hand, as can be seen, the percentage of female workers in the formal sector is higher than in the informal, this may be due to the fact that our sample excludes self-employed individuals and unpaid family workers. A much larger proportion of the workforce in the formal sector is married. In terms of the occupational structure, informal workers are most likely to be found in unskilled manufacturing and agricultural occupations (43%). Those in the formal sector are also most likely to be found in unskilled manufacturing and agricultural occupations like informal workers but at a lower rate (25%), followed by administrative staff (24%). There is little difference in the average hours of work in the two sectors. Firms with less than 3 regular employees are substantially more likely to be part of the informal sector. In contrast, larger firms employ much of the formal-sector labor force with a workforce greater than one hundred.

### **3. Wage estimates - Empirical Strategy**



Most of the former studies that intended to measure the formal – informal sector wage gap have simply estimated a Mincerian wage equation using OLS. The framework for the empirical analysis is a model in which the wage of an individual  $i$  in sector  $j$  is given by:

$$W_{ij} = \alpha_j S_{ij} + \beta_j X_{ij} + \varepsilon_{ij} \quad (1)$$

where  $W_{ij}$  denotes the log of the hourly wage of the individual  $i$  in the sector  $j$ , formal ( $F$ ) or informal ( $I$ ),  $S_{ij}$  the years of acquire education,  $X_{ij}$  denotes the set of other characteristics (e.g. experience, tenure, gender) that affect the wage of this individual; and  $\alpha_j$  and  $\beta_j$  is a vector of prices or returns associated with years of acquire education and other characteristics that affect wages.

The typical specification adopted to estimate the effect on earnings of education – occupation mismatch is based also on the Mincerian wage equation. However, the general educational mismatch specification varies slightly in that the variable of years of schooling acquire is decomposed into three variables: required, surplus and deficit education, following Duncan and Hoffman (1981) formulation. Over-education is the amount of years of schooling a worker has acquired in excess of the required education needed to perform his job. Under-education entails the opposite. Under this framework wages are a function of over, required and deficit years of education (so-called ORU wage equation). That is:

$$W_{ij} = \alpha_{rj} S_{ij}^r + \alpha_{oj} S_{ij}^o + \alpha_{uj} S_{ij}^u + \beta_j X_{ij} + v_{ij} \quad (2)$$

where  $S^r$  is years of required education,  $S^o$  is years of surplus education above the required level and  $S^u$  is years of deficit schooling below the required level. Then, under this wage equation the returns from additional education are  $\alpha_{oj}$  for surplus years,  $\alpha_{rj}$  for required years, and  $\alpha_{uj}$  for deficit years of education in sector  $j$ .

Next we want to analyze the returns to education and the effects of occupation-education mismatch on the entire wage distribution for formal and informal workers, by using linear quantile regression estimates. By estimating linear quantile regressions we are able to examine the heterogeneous effect of education at different points in the wage distribution. For any worker  $i$  in sector  $j$  we can write the  $\tau^{\text{th}}$  quantile of the hourly wage distribution conditionally on actual years of education ( $S_{ij}$ ) and other characteristics ( $X_{ij}$ ) as:

$$F_{W_{ij}}^{-1}(\tau | S_{ij}, X_{ij}) = S_{ij}\alpha_j(\tau) + X_{ij}\beta_j(\tau), \quad \forall \tau \in [0,1] \quad (3)$$

where  $F_{W_{ij}}^{-1}(\tau | S_{ij}, X_{ij})$  is the  $\tau$ th quantile of  $W_{ij}$  conditionally on  $S_{ij}$  and  $X_{ij}$ . The estimated quantile regression (QR) coefficients can be interpreted as the rates of return to actual education and other characteristics at different points of the conditional wage distribution. Similarly for any worker  $i$  in sector  $j$  we can write the  $\tau^{\text{th}}$  quantile of the hourly wage distribution conditionally on years of required education ( $S_{ij}^r$ ), years of surplus education ( $S_{ij}^o$ ), years of deficit education ( $S_{ij}^d$ ), and other characteristics ( $X_{ij}$ ) as:

$$F_{W_{ij}}^{-1}(\tau | S_{ij}^r, S_{ij}^o, S_{ij}^d, X_{ij}) = S_{ij}^r\alpha_j^r(\tau) + S_{ij}^o\alpha_j^o(\tau) + S_{ij}^d\alpha_j^d(\tau) + X_{ij}\beta_j(\tau), \quad \forall \tau \in [0,1] \quad (4)$$

However, there could be non-observable characteristics that affect wages and simultaneously affect the sector in which the individuals are currently working. This will cause to obtain not only biased, but also inconsistent coefficients. Based on this concern, a more conventional approach is to implement a selection correction in the wage regressions for each sector. This entails a two-stage estimation process. In a first stage a reduced-form probit model of the formal vs. informal decision is estimated and a sample selection correction term is obtained. In stage two, the correction term is incorporated into conventional Mincerian semi-log earnings functions for the formally employed and informally employed (see, e.g., Gong and van Soest 2002; Günther and Launov; 2012).

The selection process of the sector of employment follows the latent model:

$$E_i^* = \gamma Z_i + \mu_i \quad (5)$$

$E_i^*$  is a latent variable that determines the sector in which individual  $i$  is employed ( $j =$  formal or informal). The observed binary variable  $E_i$  is related to the latent variable  $E_i^*$  as follows:

$E_i = 1$  if the individual is in the formal sector ( $E_i^* \geq 0$ )

$E_i = 0$  otherwise

$Z_i$  is a vector of observed individual characteristics included in  $X_i$  in the wages equations plus some other variable(s) for be employed in the formal or informal sector, and  $\mu_i$  is the error term.

Estimates of returns based on the wage equations, leaving aside the selection equation (5), are biased and inconsistent if  $cov[\mu_i, \varepsilon_{ij}] = \rho_j \neq 0$ .

In the case of estimates at the mean, consistent estimates can be obtained by maximum likelihood considering the information from the two equations or, alternatively, by applying the two-step method proposed by Heckman (1979). The Heckit method includes the inverse Mills ratio in the wage equation as an additional regressor to obtain wages conditional on being in the formal or informal sector.

While the methods for correcting sample selection for mean regression are well acknowledged, there are few known approaches to correct for selectivity bias in quantile regression models and there is little consensus regarding the most appropriate correction procedure. Buchinsky (1998) suggests an approach to approximate the selection term by a power series expansion of the inverse of the Mill's ratio and is the most common approach used so far for correcting

selectivity in quantile regression models (Hyder and Reilly, 2006; de la Rica et al, 2008; Albrecht et al, 2009).

#### **4. Returns to education across sectors – empirical results**

Table 3 presents the coefficients obtained from estimating the Mincer wage equation (1) and the coefficients of estimating the ORU wage equation (2) using the mean method. The dependent variable is the gross hourly wages. Estimates were done separately for formal and informal workers. A simple specification for the two wage equations was used to account fully the effect of human capital variables. It includes as explanatory variables the number of years of education (actual years of education in the Mincerian wage equation and years of education decomposed into surplus, required and deficit in the ORU wage equation), the years of experience and its square, the months of tenure with the current firm and its square, and the gender of the individual. The results of this simple specification are presented in the first column of each estimated wage equation.

However as it has been shown in the descriptive analysis, formal and informal workers differ significantly in firm characteristics and individual characteristics, beside those related to human capital. For instance given that firms tend to be larger in the formal sector and larger firm pay more, formal workers could obtain a higher return to their education just because they are more prone to work in large firms while informal workers are more likely to work in small firms. Thus to ensure that the comparison of the returns to education across the two sectors is done for observably similar workers, a more comprehensive specification that includes additional controls was used for the two wage equations. Including additional individual and job characteristics also allow us to disentangle to what extend these observable characteristics explain the average wage differentials across formal/informal workers. Those controls include dummy variables for marital status, head of household, occupation, contract signed, size of the

firm, sector, metropolitan area and a continuous variable for hours worked. The results of this more comprehensive specification are shown in the second column of each estimated wage equation.

We start by describing the results of the Mincerian wage equation for the simple specification (column 1). The results for the total sample shows that education is better rewarded in the formal sector than in the informal sector, each additional year of schooling increase hourly wages by 10.08% for formal workers, and this is around double that for the informal workers, 5.43%. Once additional controls are accounted for (column 2) the return to schooling estimated for both sectors is lower, especially for formal workers. Each additional year of schooling increased hourly wage by 7.08% for formal workers and by 4.24% for informal workers. Nevertheless, the finding that formal workers have a higher return to their education than informal workers still holds.

Considering the existence of educational mismatches gives an interesting picture of the difference in the returns to schooling across the two sectors. Table 3 also presents the returns associated with schooling when educational mismatches are present, Eq. (2). Consistent with previous literature i) the returns to surplus schooling are lower than the returns to required schooling, ii) a year of deficit schooling carries a wage penalty for both sectors, and iii) the returns on required education are higher than that on attained education in the Mincer equation. As it can be seen, the returns to required and to surplus schooling are higher in the formal sector than in the informal. For the entire sample and with the mean method one additional year of required education raises hourly wages by 13.23% in the formal sector and by 7.63% in the informal. Years of surplus education are associated with an earning increase of 9.31% for formal workers and 4.32% for informal workers. Noteworthy is that the penalty of deficit schooling is almost the same across sectors, around 3%-4%. As with the results from estimating the Mincerian wage equation wage, when additional controls are introduced in the estimation of the ORU wage equation the returns to schooling are lower, mainly for formal workers, but

remain significantly higher than those for informal workers. To sum up, formal workers have higher returns to their years of education than informal workers, and this is so in the presence of educational mismatch.

Table 4 presents the results obtained from estimating the linear quantile regression of eq. (3) in the upper panel and eq. (4) in the lower panel using all the set of controls (dummy variables for marital status, head of household, occupation, contract signed, size of the firm, sector, metropolitan area and a continuous variable for hours worked). The results reveal important differences along the wage distribution. First the returns to actual education increase considerably along the wage distribution for formal workers. A comparable pattern is not seen for informal workers. Second the returns to surplus education behaves similarly to the returns to actual education, increases along the wage distribution for formal workers and almost constant for informal workers. Third the returns to required education increases along the wage distribution for both type of workers, but the increase is more noticeable for formal workers. Last the penalty associated to deficit education increases for formal workers while the penalty decreases for informal workers. Interesting the difference in the returns to education for formal and informal workers in the 25<sup>th</sup> quantile is minimal, while at the 75<sup>th</sup> quantile the returns to education for formal workers are around double than that for informal workers.

Figure 1 and Figure 2 plots the quantile regression results. As it can be seen the returns to education is not homogenous along the wage distribution and this heterogeneous behavior is very different for formal and informal workers.

Our estimates of the two wage equations, when taking into account that unobservable variables might influence both wages and the choice of formal/informal employment, are summarized in Table 5 for the OLS estimates. These results correspond to estimates of the wages equations augmented by a selection correction term for each sector, using the presence of children in the household and the average number of years of schooling of other household members as an instrument for assignment into the formal or informal sector. As Gunther and Launov (2012)

“we argue that selection variables should collect household-specific reasons, which influence the decision to participate in the labor market, by determining the opportunity cost of staying out of the labor market, but at the same time have no direct impact on the earning potentials of individuals”. As it can be seen, once the selectivity is corrected the returns to schooling remains higher for formal workers in the two wage equations estimated (Mincer and ORU). It is important to note that the selection term (Mills lambda) is positive and statistically significant for formal workers. This result can be interpreted as follows, a worker that has a higher probability of working in the informal sector, due to his observable characteristics, could end up working in the formal sector thanks to unobservable factors (e.g. job-search networks or ability) and gets a higher return to his education (Tannuri-Pianto et al. 2004 find a similar result for Bolivia). In the case of informal workers the selection term is insignificantly different from zero. This implies that there is no correlation between the error terms in the sectoral choice and wage equations and the estimates given in Table 3 for informal workers are unbiased.

We also re-estimate the quantile regressions of eq. (3) and eq. (4) introducing the inverse of the Mills’s ratio (and its square) following Buchinsky (1998) procedure. The results of the quantile regressions correcting for selection are presented in Table 6 and as it can be seen the results do not vary significantly once selection is account for.

## **7. Conclusions**

There is now substantial body of literature addressing the wage differential between formal and informal workers for developing countries, theoretically and empirically. In empirical analysis wage equations are estimated for each group of workers, where one of the key factors is education (and its returns). There are papers that have gone beyond the difference in the mean, finding that the wage gap is not stable along the wage distribution, estimating quantile regressions. Some works have questioned the existence of a wage gap (i.e, market segmentation) given the endogeneity caused by unobservable characteristics of the individuals, such as skills. As far as we know there is no study that considers the fact that education-

occupation mismatching is present in both formal and informal sector, and that this may be driving, at least in part, the formal/informal wage gap. In this paper we have reexamined the wage gap between formal and informal workers taking into consideration that education-occupation mismatch is present in both sectors, using the case study of Colombia.

Preliminary results for Colombia show that formal workers that possess the education required to do their job have a higher return to their education, around double, compared with their informal counterparts. Moreover these returns vary along the wage distribution and the pattern of variation along the distribution is not the same for formal and informal workers. We conclude that adding measures of educational mismatch gives important information to the analysis of the formal/informal wage gap. In particular, we show that in the informal sector not only the returns to education are lower, but the penalty that informal workers face due to educational mismatches in terms of wages are considerable higher than their formal counterparts. This evidence should be taken into consideration when assessing the issue of informality in the labor market of developing countries since it is likely to affect the allocation of skilled and unskilled workers in formal and informal jobs, and the incentives to accumulate education.

## References

Abbas, Q. 2008. "Over-education and under-education and their effects on earnings: Evidence from Pakistan, 1998–2004." *SAARC Journal of Human Resource Development* 4: 109–125.

Alba-Ramírez, A. 1993. "Mismatch in the Spanish Labor Market: Overeducation?." *The Journal of Human Resources* 28: 259-278.

Alpin, C., Shackleton, J. R., & Walsh, S. 1998. "Over- and undereducation in the UK graduate labour market." *Studies in Higher Education* 23: 17-34.



- Amaral, P. S., & Quintin, E. 2006. "A competitive model of the informal sector." *Journal of Monetary Economics* 53: 1541-1553.
- Amuedo-Dorantes, C. 2004. "Determinants and poverty implications of informal sector work in Chile." *Economic Development and Cultural Change* 52:347-368.
- Arango, L.E., Herrera, P. & Posada C.E. 2008. "El salario mínimo: aspectos generales sobre los casos de Colombia y otros países." *Ensayos sobre Política Económica* 26: 204-263.
- Attanasio, O., Goldberg, P. K., & Pavcnik, N. 2004. "Trade reforms and wage inequality in Colombia." *Journal of Development Economics* 74: 331-366.
- Badaoui, E., Strobl, E., & Walsh, F. 2008. "Is There an Informal Employment Wage Penalty? Evidence from South Africa." *Economic Development and Cultural Change* 56: 683-710.
- Battu, H., Belfield, C., & Sloane, P. 2000. "How well can we measure graduate overeducation and its effects?" *National Institute Economic Review* 171: 82-93.
- Bauer, T. K. 2002. "Educational mismatch and wages: a panel analysis." *Economics of Education Review* 21: 221-229.
- Buchel, F., & van Ham, M. 2003. "Overeducation, regional labor markets, and spatial flexibility." *Journal of Urban Economics* 53: 482-493.
- Botelho, F. & Ponczek, V. 2011. "Segmentation in the Brazilian labor market." *Economic Development and Cultural Change* 59: 437-463.
- Bosch, M. & Maloney W. F. 2010. "Comparative analysis of labor market dynamics using Markov processes: An application to informality." *Labour Economics* 17: 621-631.

Castillo, M. 2007. "Desajuste educativo por regiones en Colombia: ¿competencia por salarios o por puestos de trabajo?" *Revista Cuadernos de Economía* 26: 107–145.

Charlot, O., & Decreuse, B. 2005. "Self-selection in education with matching frictions." *Labour Economics* 12: 251-267.

Chevalier, A. 2003. "Measuring Over-education." *Economica* 70: 509-531.

Devillanova, C. 2012. "Over-education and spatial flexibility: New evidence from Italian survey data." *Papers in Regional Science*, in press.

Doeringer, P., & Piore, M. 1971. *Internal Labor Markets and Manpower Analysis*. Lexington, Mass.: Heath.

Dolton, P., & Vignoles, A. 2000. "The incidence and effects of overeducation in the U.K. graduate labour market." *Economics of Education Review* 19: 179-198.

Duncan, G. J., & Hoffman, S. D. 1981. "The incidence and wage effects of overeducation." *Economics of Education Review* 1: 75-86.

Fields, G.S. 1975. "Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs." *Journal of Development Economics* 2: 165-187.

Flórez, C. E. 2002. "The Function of the Urban Informal Sector in Employment: Evidence from Colombia 1984-2000." Documento CEDE no. 2002-04, Universidad de Los Andes, Bogotá D.C.

Gindling, T.H. 1991. "Labor Market Segmentation and the Determination

of Wages in the Public, Private-Formal, and Informal Sectors in San José, Costa Rica.” *Economic Development and Cultural Change* 39: 585.

Goldberg, P. K., & Pavcnik, N. 2005. “Trade, Wages, and the Political Economy of Trade Protection: Evidence from the Colombian Trade Reforms.” *Journal of International Economics* 66: 75-105.

Gong, X., & Van Soest, A. 2001. “Wage differentials and mobility in the urban labor market: A panel data analysis for Mexico.” IZA Discussion Paper no. 329, IZA, Bonn.

Green, F., McIntosh, S., & Vignoles, A. 2002. “The Utilization of Education and Skills: Evidence from Britain.” *The Manchester School* 70: 792-811.

Groot, W., & van den Brink, H. 2000. “Skill mismatches in the Dutch labor market.” *International Journal of Manpower* 21: 584–595.

Hartog, J., & Oosterbeek, H. 1988. “Education, allocation and earnings in the Netherlands: Overschooling?” *Economics of Education Review*, 7:185-194.

Harris, J.R., & Todaro, M.P. 1970. “Migration, unemployment and development: A two-sector analysis.” *American Economic Review* 60: 126–142.

Herrera-Idárraga, P.; López-Bazo, E.; Motellón, E. 2012. “Informality and Overeducation in the Labor Market of a Developing Country”. XREAP Working Papers 20/2012.

Hill, M. A. 1983. “Female Labor Force Participation in Developing and Developed Countries-- Consideration of the Informal Sector.” *The Review of Economics and Statistics* 65: 459-468.

International Labour Organization. 2011. "2011 Labour Overview: Latin America and the Caribbean." ILO, International Labour Office, Geneva.

Kiker, B. F., Santos, M. C., & de Oliveira, M. M. 1997. "Overeducation and undereducation: Evidence for Portugal." *Economics of Education Review* 16: 111-125.

Kugler, A., & Kugler, M. 2009. "The Labor Market Effects of Payroll Taxes in a Middle Income Country: Evidence from Colombia." *Economic Development and Cultural Change* 57: 335-358.

Leuven, E., & Oosterbeek, H. 2011. "Overeducation and Mismatch in the Labor Market." In *Handbook of the Economics of Education*, vol 4, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann 283-326. Amsterdam: Elsevier Science.

Magnac, T. 1991. "Segmented or competitive labor markets." *Econometrica* 59: 165–187.

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

Maloney, W. F. 2004. "Informality Revisited." *World Development* 32: 1159-1178.

Maloney, W. F., & Núñez, J. 2004. "Measuring the Impact of Minimum wages: Evidence from Latin America." In *Law and Employment: Lessons from Latin America and the Caribbean*, ed. James Heckman & Carmen Pagés, 109–130. Chicago: The University of Chicago Press.

Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. 1997. "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru." *Economic Development and Cultural Change* 45: 367-392.

Mavromaras, K., & McGuinness, S. 2012. "Overskilling dynamics and education pathways." *Economics of Education Review* 31: 619-628.

McGoldrick, K., & Robst, J. 1996. "Gender differences in overeducation: a test of the theory of differential overqualification." *American Economic Review* 86: 280-284

McGuinness, S. (2003). "Graduate overeducation as a sheepskin effect: Evidence from Northern Ireland." *Applied Economics* 35: 597-608.

McGuinness, S. 2006. "Overeducation in the Labour Market." *Journal of Economic Surveys* 20: 387-418.

Mehta, A., Felipe, J., Quising, P., & Camingue, S. 2011. "Overeducation "in developing economies: How can we test for it, and what does it mean?" *Economics of Education Review* 30: 1334-1347.

Mendes de Oliveira, M., Santos, M. C., & Kiker, B. F. 2000. "The role of human capital and technological change in overeducation." *Economics of Education Review* 19: 199-206.

Mondragón-Veléz, C., Peña, X., & Wills, D. 2010. "Labor Market Rigidities and Informality in Colombia." Documento CEDE no. 2010-07, Universidad de Los Andes. Bogotá D.C.

Mora, J. J. 2005. "Sobre educación en Cali (Colombia). ¿Desequilibrio temporal o permanente?: algunas ideas, 2000-2003." Documentos Laborales y Ocupacionales, no. 2 SENA, Bogotá D.C.

Núñez, J. 2002. "Empleo informal y evasión fiscal en Colombia." Archivos de Economía no. 210, Departamento Nacional de Planeación, Bogotá D.C.

Pradhan, M., & van Soest, A. 1995. "Formal and informal sector employment in urban areas of Bolivia." *Labour Economics* 2: 275-297.

Pratap, S., & Quintin, E. 2006. "Are labor markets segmented in developing countries? A semiparametric approach." *European Economic Review* 50: 1817-1841.

Quinn, M. A., & Rubb, S. 2006. "Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries." *Economics of Education Review*. 25: 147-156.

Rosen, S. 1972. "Learning and experience in the labor market." *Journal of Human Resources*, 7, 326-342.

Rumberger, R. 1987. "The impact of surplus schooling on productivity and earnings." *Journal of Human Resources* 22: 24-50.

Sicherman, N. 1991. "Overeducation in the labor market." *Journal of Labor Economics* 9: 101-122.

Sicherman, N., & Galor, O. 1990. "A theory of career mobility." *Journal of Political Economy* 98: 169-192.

Tansel, A. 2000. "Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey." In *Informal Sector*, ed. T. Bulutay. Vol. 1. Ankara: State Institute of Statistics.

Thurow, L. C. 1975 . *Generating Inequality*. New York: Basic Books.

Tinbergen, J. 1956. "On the theory of income distribution." *Weltwirtschaftliches Archiv* 77: 156-175.

Tokman, V. 1982. "Unequal Development and the Absorption of Labour: Latin America 1950–1980." *CEPAL Review* 17:121–33.

Tsang, M. C., Rumberger, R. W., & Levin, H. M. 1991. "The Impact of Surplus Schooling on Worker Productivity." *Industrial Relations: A Journal of Economy and Society* 30: 209-228.

Verdugo, R., & Verdugo, N. 1989. "The impact of surplus schooling on earnings". *Journal of Human Resources* 24: 629–643.

Table 1. Hourly wage differentials

	All		Formal		Informal		wage gap	
	Mean	SD	Mean	SD	Mean	SD	$w_F - w_I$	$w_F / w_I$
Over - educated	4627.06	3847.00	5170.34	4116.13	2379.11	1396.24	2791.23	2.17
Correct	3588.28	2747.15	4125.16	3007.49	2714.70	1602.93	1410.46	1.52
Under - educated	2665.47	1364.69	3131.68	1443.82	2366.05	1409.71	765.63	1.32
Total	3662.58	2894.68	4240.56	3193.62	2197.83	1097.70	2042.74	1.93
Observations	13797		9513		4284			

Table 2. Descriptive statistics for the main variables in the analysis

	Total		Formal		Informal	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Gross hourly wage (pesos)	3662.58	2894.68	4240.56	3193.62	2379.11	1396.24
<i>Educational Attainment</i>						
Basic Primary or below	0.14	0.34	0.09	0.28	0.25	0.43
Basic secondary	0.13	0.34	0.09	0.29	0.22	0.42
Secondary	0.37	0.48	0.38	0.48	0.36	0.48
Higher education or more	0.36	0.48	0.44	0.50	0.16	0.37
Education (years)	10.86	3.82	11.73	3.56	8.92	3.65
Age (years)	33.83	10.23	34.64	9.73	32.03	11.03
Experience (years)	17.97	11.47	17.91	11	18.11	12.45
Tenure (months)	47.75	66.21	57.7	72.7	25.67	40.93
Women	0.43	0.49	0.44	0.5	0.41	0.49
Married	0.52	0.5	0.55	0.5	0.46	0.5
Household head	0.43	0.49	0.45	0.50	0.38	0.48
<i>Occupation</i>						
Unskilled	0.31	0.46	0.26	0.44	0.43	0.5
Professionals and Technicians 1	0.07	0.25	0.09	0.28	0.02	0.13
Professionals and Technicians 2	0.04	0.2	0.05	0.22	0.03	0.18
Managers and Public Officials	0.03	0.17	0.03	0.18	0.02	0.13
Administrative Staff	0.21	0.4	0.24	0.43	0.12	0.33
Merchant and Vendor	0.16	0.37	0.15	0.36	0.18	0.39
Service Worker	0.18	0.39	0.18	0.38	0.2	0.4
<i>Type of contract</i>						
No contract	0.29	0.08	0.05	0.06	0.82	0.43
Permanent	0.48	0.5	0.65	0.48	0.1	0.3
Temporal	0.23	0.42	0.3	0.46	0.08	0.27
Hours of work (per week)	50.54	10.59	49.96	9.17	51.82	13.13
<i>Firm size</i>						
Micro (1-10 workers)	0.33	0.47	0.14	0.35	0.74	0.44
Small (11 - 50 workers)	0.2	0.4	0.21	0.41	0.16	0.37
Medium (51- 100 workers)	0.06	0.23	0.08	0.26	0.02	0.14
Large (101 workers or more)	0.42	0.49	0.57	0.49	0.08	0.27
<i>Sector</i>						
Agricultural, mining, electricity, gas and water	0.03	0.16	0.03	0.18	0.01	0.11
Industry	0.23	0.42	0.23	0.42	0.22	0.42
Construction	0.07	0.26	0.04	0.21	0.13	0.34
Sales, Hotels and Restaurants	0.29	0.45	0.24	0.43	0.41	0.49
Transportation	0.09	0.28	0.1	0.29	0.07	0.25
Financial Intermediation	0.12	0.32	0.15	0.35	0.06	0.23
Social Services	0.18	0.38	0.21	0.41	0.1	0.31
Observations	13797		9513		4284	

**Notes:** Figures are in percentages, excepting Gross hourly wage, Education, Age, Experience and Tenure whose units of measurement are indicated in parenthesis.



Table 3. Hourly wage estimates. Mincer and ORU models.

	Mincer				Mean			
	[1]		[2]		[1]		[2]	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.1008** [0.0014]	0.0543** [0.0023]	0.0675** [0.0015]	0.0378** [0.0022]	-	-	-	-
Surplus			-	-	0.0931** [0.0028]	0.0416** [0.0052]	0.0776** [0.0025]	0.0360** [0.0045]
Required			-	-	0.1323** [0.0017]	0.0763** [0.0034]	0.1191** [0.0026]	0.0663** [0.0056]
Deficit			-	-	-0.0336** [0.0035]	-0.0468** [0.0044]	-0.0282** [0.0031]	-0.0352** [0.0039]
Observations	9512	4284	9512	4284	9512	4284	9512	4284
F-statistic	1014.1	125.5	301.6	64.9	996.3	106.1	319.7	63.0
R squared (adj.)	0.39	0.15	0.56	0.37	0.46	0.16	0.58	0.38

**Notes:** [1] = experience (its square), tenure (its square) and gender are included as controls.  
 [2] = [1] + marital status, head of household, occupation, hours worked, type of contract, size of the firm, sector and region are included as controls.  
 standard errors in [].+ p<0.1, \* p<0.05, \*\* p<0.01.

Table 4. Hourly wage estimates at the mean and at various quantiles.

	OLS		QR					
			QR 25		QR 50		QR 75	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0675** [0.0015]	0.0378** [0.0022]	0.0276** [0.0010]	0.0297** [0.0031]	0.0534** [0.0016]	0.0290** [0.0019]	0.0694** [0.0026]	0.0298** [0.0018]
Surplus	0.0776** [0.0025]	0.0360** [0.0045]	0.0348** [0.0016]	0.0290** [0.0062]	0.0600** [0.0023]	0.0322** [0.0036]	0.0864** [0.0026]	0.0284** [0.0038]
Required	0.1191** [0.0026]	0.0663** [0.0056]	0.0603** [0.0018]	0.0486** [0.0077]	0.1009** [0.0024]	0.0537** [0.0045]	0.1332** [0.0028]	0.0660** [0.0047]
Deficit	-0.0282** [0.0031]	-0.0352** [0.0039]	-0.0156** [0.0021]	-0.0312** [0.0054]	-0.0213** [0.0028]	-0.0254** [0.0031]	-0.0172** [0.0030]	-0.0283** [0.0032]
N	9512	4284	9512	4284	9512	4284	9512	4284

**Notes:** standard errors in [].+ p<0.1, \* p<0.05, \*\* p<0.01.

Table 5. Hourly wage estimates. Mincer and ORU models – Correcting for selection

	Mincer				Mean			
	[1]		[2]		[1]		[2]	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0959** [0.0015]	0.0362** [0.0025]	0.0683** [0.0016]	0.0380** [0.0026]	-	-	-	-
Surplus			-	-	0.0890** [0.0029]	0.0320** [0.0053]	0.0768** [0.0026]	0.0367** [0.0048]
Required			-	-	0.1256** [0.0017]	0.0478** [0.0036]	0.1190** [0.0027]	0.0679** [0.0060]
Deficit			-	-	-0.0297** [0.0036]	-0.0313** [0.0046]	-0.0307** [0.0032]	-0.0355** [0.0042]
Mills lambda	-0.0464** [0.0144]	0.1955** [0.0129]	0.2059** [0.0431]	-0.032 [0.0562]	-0.0542** [0.0137]	0.1848** [0.0131]	0.1615** [0.0425]	-0.0502 [0.0560]
Observations	12981.00	13078.00	12981.00	13078.00	12981.00	13078.00	12981.00	13078.00

**Notes:** [1] = experience (its square), tenure (its square) and gender are included as controls.  
 [2] = [1] + marital status, head of household, occupation, hours worked, type of contract, size of the firm, sector and region are included as controls.  
 standard errors in [].+ p<0.1, \* p<0.05, \*\* p<0.01.

Table 6. Hourly wage estimates at the mean and at various quantiles – Correcting for selection

	OLS		QR					
	Formal	Informal	QR 25		QR 50		QR 75	
			Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0683** [0.0016]	0.0380** [0.0026]	0.0293** [0.0009]	0.0342** [0.0037]	0.0549** [0.0018]	0.0341** [0.0023]	0.0727** [0.0026]	0.0305** [0.0025]
Surplus	0.0768** [0.0026]	0.0368** [0.0048]	0.0362** [0.0017]	0.0355** [0.0070]	0.0587** [0.0031]	0.0373** [0.0045]	0.0885** [0.0037]	0.0293** [0.0036]
Required	0.1190** [0.0027]	0.0680** [0.0060]	0.0604** [0.0018]	0.0593** [0.0084]	0.0969** [0.0017]	0.0605** [0.0054]	0.1367** [0.0038]	0.0683** [0.0044]
Deficit	-0.0308** [0.0032]	-0.0355** [0.0042]	-0.0190** [0.0021]	-0.0336** [0.0062]	0.0049 [0.0037]	-0.0292** [0.0039]	-0.0225** [0.0042]	-0.0274** [0.0031]
N	8955	3997	8955	3997	8955	3997	8955	3997

Notes: standard errors in [].+ p<0.1, \* p<0.05, \*\* p<0.01.

Figure 1. Estimated returns to over-required- under educated years formal workers

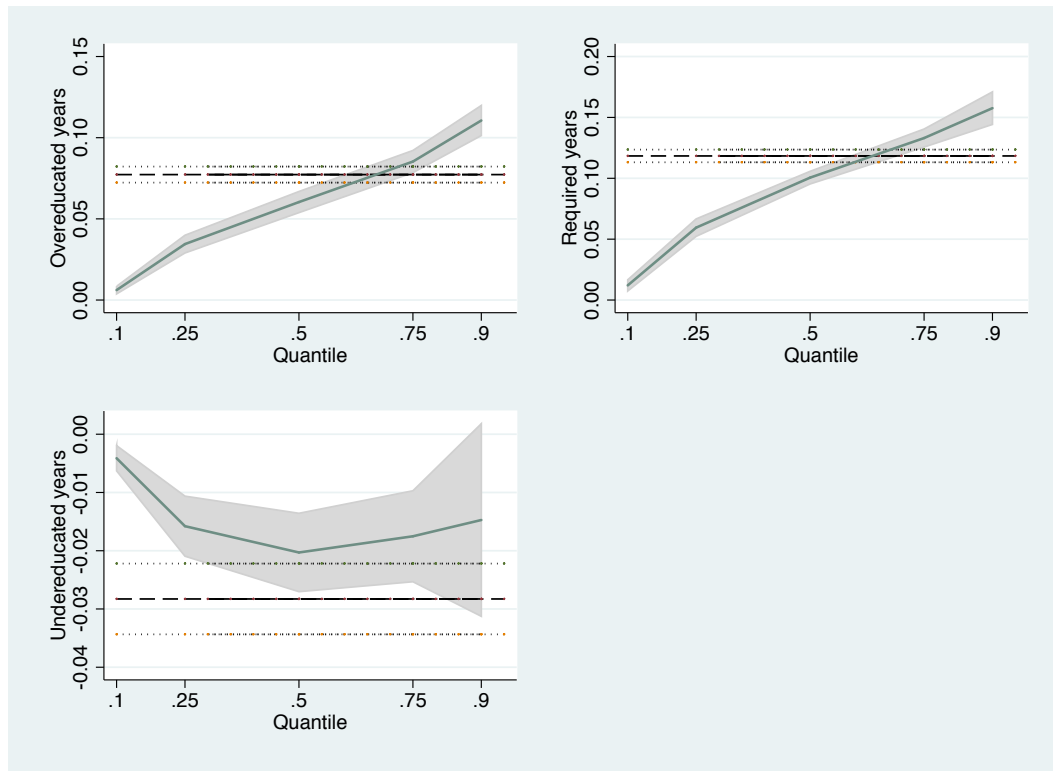


Figure 2. Estimated returns to over-required- under educated years informal workers

