Over-education and informality in the labor market of a developing country

Paula Herrera-Idárraga[a,b,*]
Enrique López-Bazo[a,c]
Elisabet Motellón[a,d]

[a] AQR – IREA, University of Barcelona, Avda Diagonal, 690, 08034 Barcelona, Spain, phone.: +34 93 4021824, fax +34 93 4021821
[b] Department of Economics, Pontificia Universidad Javeriana, Carrera 7 No. 40 – 62, Bogotá, Colombia. paulaherrera@ub.edu
[c] European Commission, Joint Research Center (*JRC*), Institute for Prospective Technological Studies (*IPTS*). Calle Inca Garcilaso 3, 41092 Sevilla, Spain. elopez@ub.edu
[d] Universitat Oberta de Catalunya, Avda Tibidabo 39-43, 08035 Barcelona, Spain. emotellon@uoc.edu

Abstract - In this paper, we explore the connection between labor market segmentation in two sectors, a modern protected formal sector and a traditional-unprotected-informal sector, and over-education in a developing country. Informality is thought to have negative implications, primarily through poorer working conditions, lack of social security, as well as low levels of productivity throughout the economy. However, in this paper we consider an aspect which has not been addressed previously, that informality could also affect the way workers match their acquired education with the one required to perform their job. We first adapt a model, which illustrates that a segmented labor market can lead to over-education in a developing economy with moderate endowments of education. Then, using micro-data from Colombia, we test the relationship between over-education and informality. Empirical results suggest that, once the endogeneity of sector choice is taken into account, formal male workers are less likely to be over-educated. Interestingly, the propensity of being over-educated for women seems not to be affected by the sector.

JEL classification*:* O15; J21; J24
Keywords: Segmented labor markets, Formal and informal sectors, Human capital, 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 7th Framework Programme project SEARCH, SSH-2010-2.2-266834.

* Corresponding author: E-mail address: paulaherrera@ub.edu

1. Introduction

In this paper, we explore the connection between labor market segmentation in two sectors, a modern protected formal sector and a traditional-unprotected-informal sector, and over-education in a developing country, Colombia. Informality is thought to have negative implications, primarily through poorer working conditions, lack of social security coverage in health and/or old age, as well as low levels of productivity throughout the economy. However, this paper considers an aspect which has not been addressed previously, which is that informality could also affect the way workers match their acquired education with the one required to perform their job. An individual worker is said to be overeducated if she has acquired more education than what it is required to perform her job. Over-education is often taken to imply that resources are wasted, because over-educated workers earn lower returns on their investment relative to similarly educated individuals whose jobs match their education. Our assumption is that over-education is not independent of market segmentation into a formal and an informal sector.

There is now a substantial body of literature that addresses the phenomenon of over-education for developed countries.[1] An increase amount of this literature is devoted to provide an explanation for over-education that is consistent with one of the theoretical labor market frameworks: Human Capital Theory (Becker, 1964), the Job Competition Model (Thurow, 1975) or the Assignment Models (Tinbergen, 1956). The majority of studies tend to support the Assignment Interpretation; within this framework the earnings will depend to some extent on both the individual and the job characteristics. These models also imply that there is no reason to expect that wage rates will be only correlated to acquired schooling or other individual attributes (Human Capital Theory), neither should be expected that individual's productivity and hence their earnings will be determined only by the job characteristics (Job Competition Model). In addition, a number of studies have also estimated the effects of over-education on earnings. These studies show that overeducated workers tend to earn higher returns to their years of schooling than co-workers who are not over-

[1]Duncan and Hoffman (1981), Verdugo and Verdugo (1989), Sicherman (1991), Tsang et al. (1991), McGoldrick and Robst (1996) studied the phenomenon for the United States; Alpin et al. (1998), Green et al. (2002), Dolton and Vignoles (2000) and Chevalier (2003) for 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 et al. (1997) and Mendes de Oliveira et al. (2000) for Portugal; Alba-Ramirez (1992) for Spain. For an extensive review of over-education in developed countries see McGuinness (2006).

educated, but lower returns than workers with similar education who work in jobs that require the level of education that they possess.

To the best of our knowledge, few studies had investigated the over-education phenomenon for developing countries. 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. One reason for the lack of studies in developing countries could be due to data limitation in order to identify the education levels required for specific jobs. On the other hand, despite the increase in the past decades in the average schooling attainment in developing countries, the average presented in these economies is less than the one presented in high-income countries. In Latin American and the Caribbean Countries the average educational attainment for those females and males between 21 and 24 years old were 9.6 years and 9.3 respectively (Duryea et al., 2007). Meanwhile the average for the OECD countries is 12.5 for males and 12.8 for females between 25 and 34 years (OECD Education at a Glance, 2010). The fact that the educational attainment remains low in developing countries makes the over-education phenomenon somewhat contradictory for these economies. Nevertheless Quinn and Rubb (2006) and Mehta et al. (2011) find evidence that over-education exists in developing countries and that the incidence of over-education is similar to that presented in developed ones.

Given the differences between labor markets in developed and developing economies, it is plausible that the explanatory factors of being over-educated may differ. For instance, labors markets of developing economies are characterized by the existence of high informality. In the particular case of Latin America, the informal sector employs between 30% and 70% of the urban work force (Maloney, 2004), embracing a variety of heterogeneous activities, such as self-employment entrepreneurs, salaried workers of large and small firms, and unpaid domestic workers. Beside the well-known negative implication of informality, primarily through poorer working conditions, a segmented labor market (in a formal and an informal sector) could also affect the way workers match their acquired education with the one required to perform their job. As Berry and Sabot (1978) affirmed, *one of the inefficiencies associated with segmentation, more difficult to document but possibly imposing greater resource costs on the economies of developing countries, involve the failure of*

the market to move the "right" resources into high wage sectors, a failure commonly described by the term "mismatch". Building on this statement, our assumption is that the study of over-education in a developing economy with a large informal sector must include as a key element the role-played by this type of segmentation.

One developing country, which is characterized for the presence of high informality in its labor market, is Colombia. Colombian informal sector is an interesting case to study for several reasons. First, informality has been one of the epicenters of the economic discussion due to the high levels prevalent in the country. According to Firm Size and Occupation criterion[2], informality was constant around 52% from 1984 to 1996, grew steadily between 1996 and 2001 to 56%, and remained at this level until 2006 (Mondragón-Vélez et al., 2010). Second, previous studies have found that in Colombia over-education exist (Mora, 2005 and Castillo, 2007). Finally the studies about informality in Colombia and other developing countries focus primarily on the size of the informal sector and on the effects of labor market rigidities on employment, wages and its distribution, and on the probability to become informal.[3] However little attention has been paid to the effects of a large informal sector in the way workers match their education with the one required to perform their job. If labor market segmentation may lead to education-occupation mismatches, then it can also affect the allocation of resources in the educational system, one manifestation of which is the bias toward academic training (Berry and Sabot, 1978).

The aim of this paper is, therefore, to extend the over-education literature for developing countries by exploring the connection between labor market segmentation and over-education. More specifically, we first adapt a model which illustrates that a segmented labor market in two sectors, a modern protected formal sector and a traditional-unprotected-informal sector, can lead to over-education in a developing economy with moderate levels of education. Then, we hypothesize that in the presence of a large informal sector in developing countries, it is possible that workers at the informal sector end up being more overeducated than formal workers in order to seek for better work conditions, after controlling for other characteristics. In the

[2] This definition is in line with the one proposed by the International Labor Office, which takes into account the occupational category of workers and the firm size, regardless of enrollment into the social security program.

context of a segmented labor market, into a formal and informal sector, while more educated workers tend to be more productive than less skilled workers, education may not provide access to good jobs. A high skill worker who is unable to get a high skill job in the formal sector may accept a low skill wage in a low skill job in the informal sector for which he or she is over-educated. Nevertheless these considerations can only constitute one of the plausible explanations of why over-education may exist in developing countries. It is known that in developing countries information about jobs and access to employers depends on the personal contacts of the individual (Tenjo, 1990). Given the importance of these informal channels, through which job search takes place in developing countries, it is probable that education mismatching occurs for those individual who don't have access to these networks.

We test the positive relationship between informality and over-education by exploiting the information in a micro-data set for Colombian workers. In doing so, two types of empirical models are used. Firstly, a simple univariate probit model that assumes that the unobservable characteristics that affect the chances an individual faces of working in either the formal or the informal sector are independent of those determining her propensity of being over-educated. Secondly, a bivariate probit model that allows controlling for the likely endogeneity of the sector in which an individual works. Results confirm that conditioned to the other individual and family characteristics, workers in the formal sector have a significant lower probability of being over-educated. This general result seems to be driven by the higher chances of being over-educated faced by male workers of the informal sectors, whereas no significant differences are detected between informal and formal female workers.

The rest of the paper is organized as follows. The next section adapts the theoretical model in Charlot and Decreuse (2005) to show that labor market segmentation in a formal and an informal sector can result in large levels of over-education in a developing country. It also serves as the framework for our empirical analysis in the following sections. Details on the data and some selected descriptive are given in section 3, while the empirical approach is presented in section 4. Section 5

[3] Magnac (1991), Nuñez (2002), Maloney and Nuñez (2004), Floréz (2002), Kugler and Kugler (2009) and Mondragón-Vélez et al. (2010) for Colombia; Pradhan and van Soest for Bolivia (1995); Pratap and Quintin for Argentina (2006); Tansel (1999) for Turkey and Gong and Van Soest (2002) for Mexico.

summarizes the results regarding the estimates of the empirical models, section 6 presents some robustness checks and, finally, section 7 concludes.

2. Theoretical framework: search frictions in the labor market, labor market segmentation and over-education

Charlot and Decreuse (2005) showed that over-education takes place when labor market is segmented and heterogeneous workers allocate themselves between sectors depending on their decision to invest in education. Here, we adapt Charlot and Decreuse (2005) model to illustrate that a segmented labor market in two urban sectors, a modern protected formal sector and a traditional-unprotected-informal sector, can lead to over-education. In our opinion, this is a reasonable explanation for educational mismatching in labor markets of developing countries. In contrast with (some) developed countries in which over-education is clearly associated with large endowments of education, the population in developing economies shows low or moderate levels of education attainment. Segmentation in the formal and the informal sector is, thus, a phenomenon that could be behind over-education in these economies.

The model in Charlot and Decreuse (2005) is based on three main figures that we adjust here for the case of an informal and a formal sector. First, workers can direct their search into the formal or informal sector, second the formal sector is intensive in high skilled workers and the informal sector is intensive in low skilled workers. Third, workers are heterogeneous; there are two schooling levels, high and low. Variables are indexed by k and j, where $k = h$ stands for high and $k = l$ for low and $j = f$ stands for formal sector and $j = i$ for informal sector. The output y_j of a match depends on the worker's characteristic a (ability) and on the sector-specific productivity parameter A_j (technology). The technology currently in use in the formal sector is more efficient than that used in the informal sector $(A_f > A_i)$. The productivity of formal firms could be higher than that of informal firms because a higher capital-labor ratio, caused by the fact that informal firms may have less access to credit, and choose to substitute low-skill labor for physical capital (Amaral and Quintin, 2006). Another reason is that informal firms continue to operate at a small size that allows them to scape from government control and, therefore, cannot exploit possible economies of scale.

Vacant jobs and unemployed workers are brought together in pairs through an imperfect matching process. The total number of employer-worker contacts M_j on sector j is given by the following matching technology:

$$M_j \equiv M(u_j, v_j) \qquad (1)$$

where u_j is the number of unemployed and v_j denote the numbers of vacancies on sector j. The function $M(u_j,v_j)$ is assumed to be twice differentiable, increasing in its arguments and to exhibit constant returns to scale. The labor market tightness is determined by the ratio between the vacancy rate and the number of unemployed ($\theta_j \equiv v_j/u_j$) for each sector. Because of constant returns to scale, it is possible to write the rate at which vacancies are filled on sector j as $M_j/v_j \equiv m(\theta_j)/\theta_j$, a decreasing function of θ_j. Similarly, the unemployed meet jobs in a sector j at rate $M_j/u_j \equiv m(\theta_j)$, an increasing function of θ_j.

If it is assumed, for simplicity, no job quits and no unemployment benefits, then job seekers move out of unemployment with probability $m(\theta_j)$ and enjoy a gain from state change equal to the difference between the asset value of being employed, $W_j(a)$, and of being unemployed, $U_j(a)$. On the other hand, firms gains are a function of the asset values of holding a vacancy, V_j, and a filled job, $J_i(a)$, in sector j. It is assumed that holding a vacancy induces a flow cost $\gamma>0$. Vacancies may become filled with probability $m(\theta_j)/\theta_j$ and bring a revenue to the firm equal to $y_j(a) - w_j(a)$, where $w_j(a)$ denotes the wage of a worker endowed with ability a after discounting for the effective rate r.

There is a rent sharing process between workers and firms over the surplus of the match, which is modeled using a standard Nash bargaining framework. The wage is a positive function of the labor market tightness, the output flow, and the bargaining power of the workers in each sector. If the output flow increases then the size of the surplus that can be shared between firms and workers rises and firms are able to pay higher wages. Likewise, a rise in the bargaining power of workers and a higher ratio of vacancies to unemployed workers, i.e. higher tightness, raises the share of the surplus that workers can convey.

Heterogeneity of the labor force and schooling costs gives rise to self-selection in the educational system, another key assumption in the model. Under this assumption agents face a binary choice, i.e. being educated or uneducated. Let $C > 0$ denote the (ability-independent) schooling cost. At birth, each person compares his/her utility if educated (net of schooling costs) to that he/she would get if uneducated and invest accordingly. An individual endowed with characteristic *a* decides to get an education if and only if:

$$U_h(a) \geq U_l(a) + C \tag{2}$$

The utility levels depend on expected wages and also on the job-finding rate $m(\theta_j)$ in each sector. As workers are charged a fixed cost for their schooling investment, only those with sufficiently high ability (*a*) choose to invest in education and are qualified to work on high–productivity jobs. The average abilities across the pools of unemployed in each sector are functions of an endogenous cutoff point, σ, below which individuals do not acquire education, and satisfy:

$$E_l(\sigma) = \int_0^{\sigma} \frac{\phi(a)}{\Phi(\sigma)} ada \quad \text{and} \quad E_h = \int_{\sigma}^{1} \frac{\phi(a)}{1-\Phi(\sigma)} ada \tag{3}$$

If educational attainment is positively correlated with labor market outcomes (wages and employment), self-selection gives rise to a composition effect. According to this effect, the mean ability among each sector decreases when the share of educated rises, i.e when the selection threshold, σ, decreases. As only the ablest get an education, the number of educated rises only if less-able individuals are drawn into education: the ablest among former uneducated now become the least able among educated. As a result, the average productivity across the sectors falls. By differentiating equation (3) we see this effect of the threshold, σ, on the average abilities in each sector:

$$\frac{dE_l(\sigma)}{d\sigma} = \frac{\phi(\sigma)}{\Phi(\sigma)}\left[E_l(\sigma) - \sigma\right] > 0 \tag{4}$$

$$\frac{dE_h(\sigma)}{d\sigma} = \frac{\phi(\sigma)}{1-\Phi(\sigma)}\left[\sigma - E_h(\sigma)\right] > 0 \qquad (5)$$

Self-selection into education creates a composition effect: the mean ability among each sector decreases when the share of education rises. This composition effect interacts with the firms' incentives to enter each sector and an externality arises. The tightness of each market is a function of the productivity, A_j, and of the average ability across the pools of unemployed, $E_j(\sigma)$, in each sector. The lower the sector-specific mean ability the lower is the firms expected profits. The lower profits attract fewer firms in the industry and fewer vacancies are advertised. The drop in the labor market tightness increases the return to education, since it is relatively more important to raise one's job-finding rate when the market is loose. Workers acquire some excess education to improve their opportunities during the wage bargain and raising their chance of leaving unemployment. This leads to over-education, since the threshold individual considers her own earnings and employment prospects, which improve by schooling, but does not internalize the impact of her schooling decision on others' job opportunities, too many workers are willing to educate and over-education arise.

The adaptation of the Charlot and Decreuse (2005) model to a labor market segmented in a formal sector and an informal one can, thus, been used to explain the existence of over-education even in an economy in which educational attainment of the population is still low or moderate, such as the Colombian one. However, it does not predict in which sector the incidence of over-education will be higher. In the following sections, our empirical exercise tries to shed light on this issue. After showing evidence on the existence of educational mismatching in Colombia, distinguishing between that observed in the formal and in the informal sector, we estimate the impact of informality on the probability of a Colombian worker to be over-educated, conditioned to the effect of the individual and household characteristics.

3. Data and descriptive statistics

We use data from the 2008 wave of the Colombian Household Survey (CHS), a repeated cross-section conducted by the National Statistics Department (DANE). This

survey gathers information about employment conditions for population aged 12 years or more including income, occupation, industry, and firm's size, in addition to the general individual's characteristics such as sex, age, marital status and educational attainment. Some household characteristics, such as the head of the household, the number of children, and the level of education of all its members, are also available. The CHS covers the thirteen major metropolitan areas in Colombia that accounted for 45% of the country's population in 2008. It must be mentioned that this survey has been used for empirical studies analyzing labor market issues in Colombia in other previous studies (Magnac, 1991; Attanasio et al., 2004 and Goldberg and Pavcnik, 2005).

The analysis in this paper was restricted to employed individuals aged between 15 and 60 years that were not carrying formal studies and reported working between 16 and 84 hours per week. Government employees, household employees, self-employed, bosses or employers, unpaid family workers, workers without pay in enterprises or other families business and journeyman or pawn were not included in the sample. The size of the sample used in the analysis was of 15104 observations.

As a starting point in our analysis, we had to use a criterion to determine if a worker in the sample was over-educated, and if she worked either in the formal or in the informal sector. Four basic methods had been suggested in the literature for measuring the required education for a job and consequently over-education. The first 'subjective' approach uses self-assessment to define the job's educational requirements and then compares this with the worker's acquired education (Battu et al., 2000 and McGuinness, 2003). The second is a variation of the previous and consists in asking directly the worker whether or not she or he is over-educated (Devillanova, 2012). Over-education can also be calculated objectively by using job analyst definition of the educational requirement for each occupation, as available in the United States Dictionary of Occupational Titles, and comparing this with the educational level of the worker (Rumberger, 1987; Hartog and Oosterbeek, 1988; Kiker and Santos, 1991; Chevalier; 2003). Another objective measure is obtained by analyzing the distribution of education for each occupation; employees who depart from the mean (Verdugo and Verdugo 1989) or mode (Mendes de Oliviera et al.

2000) by more than one standard deviation are classified as over-educated. This last approach is usually known as the 'statistical' method. .

Since the CHS does not supply information to construct a subjective measure of over-education, and taking into account that the requirements of education in the rather broad categories of occupations (two-digit ISCO classification) available in the CHS are likely to differ from that in the US economy, we decided to follow other studies in the literature in applying the statistical approach based on the mean of the distribution of education within each two-digit occupation.[4] Using such an objective measure, the overall incidence of over-education in the sample is 15%, similar than that reported for other developing economies (Quinn and Rubb, 2006) and lower than the incidence of over-education in the developed economies (McGuinness, 2006).

Regarding informality, alternative definitions and corresponding ways of measuring it have been proposed in the literature. The lack of agreement on how to define and measure it has mostly to do with the availability of data in each study. In the particular case of Colombia, according to DANE, informal workers are those who fall into at least one of these categories: (i) work in firms with five or fewer employees; (ii) are unpaid family workers and housekeepers; (iii) are self employed who work in places up to five persons other than independent professionals and technicians; or (iv) are business owners of firms with five or less employees. This definition, in line with the one proposed by the International Labor Office, has been criticized in the literature because it does not take into account the benefits associated with formal employment, such as the inclusion in the social security system (Flórez, 2000). Nevertheless, the available data from the CHS permits us to determine whether workers in the sample are covered or not by the social security system, and even distinguish between contributions to the retirement pension and to the health system. Using this information, we classified workers as formally or informally employed according to their full inclusion in the social security system. That is to say, an individual was classified as a formal worker if she contributed to both health and

[4] As stressed in Ramos and Sanromà (2011), a 2 digit classification of occupations is not optimal for applying the mode criterion. In any case, we also computed the results in the following sections using this criterion as a robustness check, and the main conclusions remained the same as those derived from results using the mean criteria.

retirement pension systems. Under such condition, as much as 33.3% of individuals in the entire sample worked in the informal sector.[5]

The incidence of over-education in the entire sample and in the two sectors, and the percentage of Colombian workers in the formal and in the informal sectors are shown in the first group of rows in Table 1. This table also displays basic summary statistics concerning the distribution of individual's and job's characteristics considered in the subsequent analysis, distinguishing between those for workers in the formal and in the informal sector. In addition, to get some insight into the existence of differences by gender in the magnitudes under analysis, figures have been computed for the entire sample and for men and women.

As already mentioned above, 15% of Colombian urban workers were over-educated, this figure being higher in the case of the formal workers (17%) than for those in the informal sector (11%). This gap of 6 percentage points exists for both male and female workers. Regarding the distribution of workers in each sector, around one third had a job in the informal sector in 2008, this percentage being higher for men (35%) than for women (31%). Differences in over-education between the two sectors, and by gender, might be simply caused by disparities in the distribution of the characteristics that are assumed to affect the incidence of over-education. Actually, the comparison of figures in Table 1 confirms that there are substantial differences in some of the observable worker and job characteristics between formal and informal workers. As a matter of example, the years of schooling, as a measure of education, are not only useful as a proxy for general human capital but they are also likely to be correlated with unobserved individual's ability. What the figures show is that workers in the informal sector are more likely to have education levels below that of those employed in the formal sector: whereas 45% of informal workers in the entire sample have at most basic secondary education, the percentage of workers in the formal sector with secondary or tertiary education is as high as 81% (45% with tertiary education). If, as expected, there is a strong association between education and the likelihood of over-education, such a gap in educational attainment could explain a big deal of the difference observed in the over-education figures between the two sectors.

[5] Self-employment in Latin America generally constitutes one of the principle sources of employment and a large of it operates in the informal sector, if the sample is not restricted to exclude self-employees the percentage of informal workers increased up to

There are significant differences in other characteristics as well. The percentage of female workers in the formal sector is higher than in the informal, perhaps due to the fact that our sample excludes self-employed individuals. A much larger proportion of the workforce in the formal sector is married, and workers in that sector tend to accumulate much more tenure than informal workers, suggesting a higher stability of employment in the formal sector. As for the occupational structure, the share of informal workers in unskilled manufacturing and agricultural occupations (42%), and in merchant, vendor and service worker (36%) is larger than in the formal sector, while the administrative staff (24%) and professionals and technicians are more represented in the formal sector. Finally, it is worthwhile mentioning that more than two thirds of informal workers are employed in small firms, with 10 or less workers. This is in sharp contrast with figures on the distribution of workers by firm size in the formal sector, as more than half of them work in firms with more than 100 employees, and around two thirds in firms with at least 50 employees. In a nutshell, these figures indicate that there is a clear connection between informality and firm size in Colombia.

As for the gender component, Table 1 shows that male and female workers differ in some of the characteristics that are supposed to affect over-education. Interestingly, the most remarkable differences appear in the distribution of levels of education and in occupations. Broadly speaking, female workers are more educated than their male counterparts, and are concentrated in occupations such as administrative staff (24%), merchant and vendor (22%) and service worker (20%), while men are more concentrated in unskilled manufacturing and agricultural occupations (48%), that are associated with higher levels of informality.

This simple descriptive analysis suggests i) the existence of rather large levels of over-education in Colombia, ii) that seem to affect more intensively workers in the formal than in the informal sector, and iii) that formal and informal workers differ in their levels of educational attainment, the occupational distribution, and other individual and job characteristics that are though to exert an influence on the

59% for 2008.

individual's probability of over-education. Since the higher incidence of over-education in the formal sector might well be caused by a composition effect (for example associated to the higher education of workers in that sector) in the next section we estimate the impact of informality on over-education conditioned to the set of observable worker and job characteristics.

4. Informality and over-education. Empirical specification

A multivariate empirical model needs to be specified in order to assess the impact of the sector on the Colombian workers' probability of being over-educated, conditioned to the other observed individual, household and job characteristics. In doing so, in a first step we assume that the allocation of a worker in a formal or an informal job is exogenous to her chances of being over-educated. Under such assumption, a univariate probabilistic specification provides consistent estimates of the effect of the sector on the chances of having more education than that required for the worker's occupation. However, the endogeneity assumption can easily be questioned. For instance, it can be argued that unobserved individual ability might well be affecting both the propensity to work for the formal sector and to be correctly matched. In such a case, the standard probabilistic specification with exogenous covariates lacks consistency. To address this issue, we estimate the effect of the sector by means of a bivariate specification in which the variable of interest is instrumented.

In brief, a simple way to identify the determinants of educational mismatch is to assume a latent continuous (unobserved) variable Y_i^* for the probability of over-education of worker *i*, which is related to a linear index function and an additive error term, ε_i :

$$Y_i^* = \beta X_i + \alpha S_i + \varepsilon_i \qquad (6)$$

where X_i is a vector of individual and firm characteristics (such as age, gender, marital status, head of household, education, tenure, occupation, industry sector, contract type and the unemployment rate of the metropolitan area), S_i is a dummy

variable for the sector, formal or informal, and ε_i is a normally distributed error with zero mean and unit variance.

The observed dichotomous realization Y_i of the latent variable Y_i^* is as follows:

$Y_i = 1$ if the individual is over-educated ($Y_i^* \geq 0$)

$Y_i = 0$ otherwise

Given the normality of the error term in (6) a probit specification can be used to estimate the effect of the sector on the probability of being over-educated, conditioned to the other characteristics in *X*:

$$Y_i = \Phi(\beta X_i + \alpha S_i) \tag{7}$$

Since the estimate of the coefficient α is only informative about the sign of the impact of *S*, its associated (average) marginal effect is computed from the estimates of the probit model in (7) as:

$$\partial\left[P(Y=1)/S\right]_{\bar{X}} = \Phi(\beta \bar{X} + \alpha) - \Phi(\beta \bar{X}) \tag{8}$$

where the bar over the X denotes the sample average.

Likewise from the estimates of the probit model it is also possible to calculate the Average Treatment Effect (ATE) as:

$$E[\delta P(Y_i = 1)/\delta S_i] = (1/n)[\Phi(\beta X_i + \alpha) - \Phi(\beta X_i)] \tag{9}$$

where *n* denotes the number of individuals in the sample.

As indicated above, the assumption that is made in the specification of the univariate probit in (7) is that the sector of employment, formal or informal, is exogenous to the probability of being over-educated. However, if the assignment of workers in each of the sectors is not random and some unobservable factors - ability among others - that

influence the probability to be assigned into a particular sector of employment could also affect the probability of being overeducated, then the estimation of a univariate probit suffers from selection bias.[6] This has dramatic consequences on the inference since the estimates from the univariate probit are inconsistent if such a type of endogeneity is ignored.

To properly take account of this drawback, in a second step we estimate the effect of the sector of employment in a bivariate probit model, in which the sector is instrumented by family characteristics. In addition to the outcome latent equation in (6), the bivariate model is based on an additional equation for the latent model linking the probability of assignment in the formal or in the informal sector to a set of characteristics:

$$S_i^* = \gamma Z_i + \mu_i \qquad (10)$$

where Z_i is a vector of observed individual and family characteristics, and μ_i is the error term. Z_i includes the set of characteristics in X_i plus some additional variables used as instruments for the sector of employment, S_i.

Since we can only observe the sector of employment for each individual, the link between the observed binary variable S_i and the latent variable S_i^* is assumed to be as follows:

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

$S_i = 0$ otherwise

Therefore, the probit specification associated to the probability of working in the formal sector conditioned to the characteristics in *Z* stands as:

$$S_i = \Phi(\gamma Z_i) \qquad (11)$$

[6] We left aside another type of selectivity concerning the fact that an individual might not accept a job that does not match his or her level education and chooses instead to be unemployed or outside the labor force. We argue that this selectivity is not relevant in the Colombian case where there is no unemployment benefit system and the family protection network against unemployment is low or exclusive for some group of individual with high income.

The bivariate probit thus consists of equations (7) and (11), where μ_i and ε_i are distributed bivariate normal, with $E[\mu_i] = E[\varepsilon_i] = 0$, $var[\mu_i] = var[\varepsilon_i] = 1$ and $cov[\mu_i, \varepsilon_i] = \rho$. In other word, the empirical model allows for the likely correlation of the unobserved determinants of over-education and the unobserved determinants of the sector of employment.[7] In such a framework, there are four possible states of the world (Y_i =0 or 1 and S_i = 0 or 1), and the corresponding log-likelihood function associated to this set of events is (for further details see Wooldridge (2002) page 478, 2002).

The inference in the bivariate probit model is based on maximization of the log-likelihood in eq. (12) with respect to the parameters β, α, γ and ρ. If ρ is statistically different from 0, estimates from the bivariate probit are preferable; otherwise conclusions on the impact of the sector of employment could be based on the estimate of the univariate probit in eq. (7). Marginal effect and Average Treatment Effects are computed from the estimates of the biprobit model using a similar formulation as for the univariate probit model.

Two issues that usually result from the estimation of a bivariate probit model with an endogenous binary regressor are identification and the selection of valid instruments. Identification can be achieved by relaying sorely on the functional form and distributional assumptions. However, the objective of forming a consistent estimator for α becomes manageable if one can construct at least an instrument for S_i . A variable I_i would be a valid instrument for S_i if it were a determinant of the sector of employment and it were not correlated with the error term of the over-education equation (outcome equation). The first condition is easy to meet; we can verify whether Z_i is correlated with S_i , once the other variables have been controlled for. However, it is harder to test if the instrument is valid or not. This condition relies on the economic or institutional knowledge concerning the problem under study.

As it is the case in many other studies, finding suitable instrumental variables is far from straightforward, since almost any regressor that determines the probability of

[7] For instance, a bivariate probit model with an endogenous binary regressor has been used in Evans and Schwab (1995) to

being over-educated could plausibly affects the assignment in the sector of employment as well. Still, it is our belief that some family characteristics influence individual's choice of sector of employment but not affect over-education in a direct manner. One of such characteristics is the presence of children in the household. The assumption here is that the presence of children does not exert a significant effect on the propensity of being over-educated. It could be argued that some individuals may be willing to accept such educational mismatching in exchange for other labor conditions that may compensate them. For example, having more flexibility in working hours that allows them to take care of children. Whereas this may be the case in modern societies for developed countries[8], it hardly occurs in more traditional societies of developing countries. In developing countries, like Colombia, gender roles and stereotypes still prevail; woman role is to take care of children, while the role for man is to be the breadwinners. Thus, we assume that the number of children affect the sector of employment but not the probability of being over-educated, especially for men in developing countries. Another family characteristic that is thought to influence the choice of sector of employment but not the individual's propensity of over-education is the social status, which we suggest it is capture by the educational achievement of other members of the household. Accordingly, we construct the average years of schooling of other members of the household and use it as an additional instrument for the sector of employment.

5. Informality and over-education results

Maximum likelihood estimates of the coefficients of the univariate probit model are reported in Table 2 for the entire sample, and for male and female workers. The corresponding marginal effects for the average individual and the average treatment effects are also reported. Given that both effects are similar, the discussion will be center on the marginal effects. The results show that after controlling for other characteristics, formal workers have a lower probability of being over-educated than informal workers. That is to say, when we compare formal and informal workers with similar individual, household, and firm characteristics, workers in the former group have a lower propensity of being over-educated than those in the latter group. This is

analyze the effects of public and catholic schools on finishing high school and starting college.

[8] As far as we know, Mavromaras and McGuinness (2012) is the only study that uses the presence of children as a control variable in probits estimates of over-skilling. They obtain only a marginal statistical significance for the coefficient of this variable, and just for the group of moderately over-skilled workers.

in sharp contrast with the raw probabilities derived from the sample since, as shown in the descriptive analysis in section 3, the share of over-educated workers in the formal sector is higher than the one in the informal sector. Results, thus, suggest that of sorting effect drives the gap in the raw propensities.

In any case, it should be mentioned that the marginal effect associated to working in the formal sector is of a moderate magnitude. The probability of being over-educated for a formal worker is just 2.5 percentage points (pp) less than that for a similar worker in the informal sector. The impact on the probability is even lower for men, 1.86 pp, being relatively higher for women, 2.72pp. Interestingly, the coefficient and the corresponding marginal effect are statistically significant only at 5% in the separate samples for both genders. Therefore, results from the univariate probit model suggest an almost negligible impact of formality on over-education once controlled for education and other observable characteristics. However, it must be kept in mind that such a specification assumes exogeneity of the sector of employment and the lack of a simultaneous impact of the unobservable characteristics on the probability of over-education and on the sector assignment. Violation of these assumptions would invalidate the results.

As for the estimate of the coefficients for the control variables, results in Table 2 are consistent to what is found in the previous literature. For the sake of brevity we only discuss next the results for the total sample. As expected, the probability of being over-educated increases with educational attainment (Alba-Ramirez, 1993; Kiker et al., 1997 and Quinn and Rubb, 2006). Overeducated workers may substitute education for the lack of previous job experience, taking jobs that required less education than they actually acquire in order to accumulate experience and improve their chances of finding a better job match (Rosen, 1972; Sicherman and Galor, 1990 and Mendes de Oliveira et al., 2000). Therefore we expect that over-educated workers will have lower experience. To test this hypothesis we use a variable that measures experience, particularly potential experience calculated as an individual's age minus five minus years of education. On the other hand, several studies have established that over-education may have a negative effect on job satisfaction (Tsang et al., 1991), if this is the case, then overeducated workers with more tenure in a firm are expected to be more prone to turnover. Consequently we hypothesize that over-educated workers

will have less tenure. The results for the estimated marginal effect of general experience confirm the expected negative effect of this variable on the probability of being overeducated for an average worker in the sample. However, it must be mentioned that such marginal effect is only significantly different from zero at 10%. The impact of tenure is also negative, though almost negligible and, actually not statistically significant. Therefore, these results for Colombia are in conflict with the evidence on the substitutability between education and other forms of human capital postulated by the human capital theory, under which over-education might be seen as a transitory situation.

The results also indicate that females are less likely to be overeducated than males with similar characteristics, and that marital status does not have a statistically significant impact on the probability of being over-educated for both genders. Significant differences exist regarding the industry and firm size. With respect to individuals working for Agriculture, mining, electricity, gas and water (reference category), workers in construction are more likely to be over-educated, whereas workers in transportation, financial intermediation and social services are less prone to be over-educated. As for the size of the firm, the incidence of over-education conditioned to the other characteristics is slightly lower for small (4 to 50 workers) and substantially lower for medium size firms (51 to 100 workers). Finally, it is worthwhile mentioning that local labor market conditions seem to not be relevant, as the coefficient of the metropolitan unemployment rate, although positive, is not statistically significant.

Results for the estimation of the effect of the sector of employment relaxing the assumption of exogeneity and lack of correlation between the unobservables that influence both over-education and formality/informality are summarized in Table 3. They correspond to the maximum likelihood estimates of the bivariate probit model described in section 4, using instruments for the sector of employment and the same set of control variables as those in the univariate probit model. We just focus the discussion here on the coefficients of the equation for the probability of being over-educated since the estimates obtained for the parameters in the formal/informal sector equation are relatively standard, and largely conform to results reported elsewhere (Magnac, 1991 and Pradhan and van Soest, 1995). They are reproduced in Table A1

in the appendix. We first make comments regarding the entire sample of Colombian workers, and then discuss some differences observed between the samples of male and female workers.

The coefficient of the formal sector and the corresponding marginal effect are estimated to be negative and highly significant. Actually, the magnitude of the marginal effect of working in the formal sector estimated from the biprobit model is substantially higher than the one obtained in the univariate probit model. The results suggest that, for otherwise similar workers, being in the formal sector reduces the probability of over-education by 15.03pp. The average treatment effect (ATE) calculated is higher as well, 15.91pp. This evidence confirms that selection bias strongly affects the estimate of the effect of the sector of employment on the probability of become over-educated, and thus the importance to account for it. On the other hand, it seems that in addition to the benefits related with social security and higher wages being a formal worker also ensures a better use of her skills in her work. Seen from the other side of the coin, informal workers, in addition to receiving lower wages and have no health and pension coverage, are less likely to properly use their knowledge acquired into their work. As we mentioned in the introductory section, we are not aware that this fact was shown in the past, being therefore a novel contribution of this piece of research.

On the other hand, the estimate of ρ (correlation between the error terms of the over-education and the employment sector equations) is positive and statistically significant. This finding suggests that non-observable characteristics that exert a positive effect on the probability of being employed in the formal sector also have a positive impact on the probability of being over-educated. This could be interpreted as evidence that for formal workers over-education is caused, at least to some extent, by the desire to become part of the formal sector (better employment opportunities, social system protection, etc.). A worker with considerable education possibly will enter a job for which less education is required, because that job is protected, for example, by the minimum wage, while the occupation for which the worker is best suited pays less than the minimum in the informal sector.

Another interpretation of the positive effect of unobservable factors over the probability of being over-educated can be understood within the internal labor market framework (Doeringer and Piore, 1972). Internal labor markets are those where workers are hired into entry-level jobs and higher levels are filled from within. Some rules differentiate members of the internal labor market from outsiders and accord them rights and privileges that would not otherwise be available. Typically these internal rights include certain guarantees of job security and opportunities for career mobility. If an internal labor market exists, then there must be some jobs, presumably at high levels, that are filled almost exclusively through internal promotion and there must be other ports-of-entry jobs, presumably at low levels, that are filled through external hiring. In this context individuals at any given firm are hired into lower or middle levels of the firm and then achieve to advance to higher levels. Workers that do not have the qualifications for particular entry-level jobs are excluded from access to the entire job ladder. For that reason workers may accept at the beginning a job in which their education acquired is higher than the required for the job, in exchange for the benefits of being part of an internal labor market. It is important to notice that internal labor markets operate in the primary sector (formal), rather than in the secondary sector (informal).

However, if ability is considered a non-observable characteristic, and according to the theoretical framework presented before, a positive correlation term of the errors seems somewhat contradictory. One will expect that workers with more ability, and hence with higher skills, end up working in the formal sector, and that this ability prevents them from being over-educated. In this case ρ will be negative rather than positive. Our results imply, therefore, that this negative effect of ability is more than offset by the positive effect of the desire to become part of the formal sector, among other factors.

As for the estimate of the coefficients, and the associated marginal effects and ATEs, of the other observable characteristics in the over-education equation, it can be said that, in general, they are roughly identical to those estimated with the univariate probit, except for the size of the firm. The estimates from the biprobit model indicate that when compared with individuals working for micro-firms (those with less than 10

workers), workers in small, medium and large firms has higher chances to be over-educated. This result might be interpreted as follows; large firms usually have better job opportunities, aside of higher wages, in this type of firms workers have more probability to be promoted and more training on the job. These characteristics make job's offers from large firm high valuable for job seekers who compete for this works. In this competition, workers could end up applying to vacancies in which the required education is less than the acquired. Likewise, employers of large firms in the formal sector are in a position to select the most skilled from the pool of available workers. In any case it must be mentioned that the impact on over-education is weaker in the case of medium size firms (between 50 and 100 workers), for which the coefficient is in fact not statistically significant.

Finally, it is important to stress that the results by gender point to a substantial difference in the impact of the sector of employment on the probability of being over-educated. Whereas working for the formal sector for a male reduces the propensity of over-education in more than 20pp with respect a an otherwise similar male worker in the informal sector, there not seems to be any significant difference in the case of females (the corresponding marginal effect and the ATE is not statistically significant). Interestingly, in the group of female workers there is not a significant correlation between the errors of the two equations either. In contrast with the highly significant correlation coefficient for males, the lack of correlation for women indicate that the unobservable characteristics that affect their chances of working in the formal or in the informal sector are independent of those affecting their propensity to be overeducated. The arguments mentioned above on the different social roles of men and women in a society such as the Colombian one might be behind such outstanding gender differences in the link between informality and over-education.

In any case, it needs to be argued that the instruments might be weak in the case of women; the number of children may affect the sector of employment and the probability of being over-educated. Additionally it must be kept in mind that we have left apart another type of selection bias that has been usually taken into account in the over-education literature; over-educated workers are a selected group of employed individuals since it is not possible to observed the over-education of unemployed. This type of selection mechanism may be more relevant in the case of women.

However we do not know any method that can correct this double endogeneity when a discrete outcome is analyzed. Finally, it must also be remembered that our sample excludes self-employed individuals and women constitute a large fraction of informal self-employment.

6. Robustness Checks

As in other past studies about over-education, our results can be sensitive to the method employed to measure the phenomenon. This issue is especially important in studies that inquire what are the determinants of over-education at the micro level. As a robustness check, we replicate all the estimations using the mode criteria. We calculate an incidence of over-education of 28.96% using the mode method; almost twice to what we obtain using the mean method.[9] As in the case for the mean method, the incidence for formal workers is higher. The incidence for formal workers is 33.05% while that for informal workers is 20.72%, which represents a difference of 12.33 pp.

Here we briefly described the differences in the effect of the employment sector has on the probability of being over-educated obtained between the two methods. Estimates of the coefficients of the univariate probit model using the mode method are reported in Table 4. The effect of the sector of employment on the probability of being over-educated for the total sample measure with the mode is very similar to that reported in Table 3 using the mean method. For men the coefficient using both methods are quite the same, both are negative, but with the mode is statistical significant at 5% and the magnitude of the marginal and of the average treatment effects are lager, -5.13pp and -4.79pp respectively. For women the effect of being in the formal sector changes of sign. However is not statistical significant and the size of the marginal and average treatment effects are low, 1.35pp and 1.14pp respectively.

Results for the estimation of the biprobit model with the mode measurement are presented in Table 5. For the total sample and for men the results are largely conform to those obtain with the mean method reported in Table 3. Regarding women the

results are different, not only in magnitude but also in sign and statistical significance. Formal sector variable for women using the mode method is positive, 19.97pp, but significant only at 10%. However the correlation between the error terms of the over-education and the employment sector equations is not statistical significant. In the study conducted by Verhaest and Omey (2010) they found that different measures of over-education could cause opposite results for women.

Our results appear to be robust to the method used for measuring the incidence of over-education, given that the signs and significance levels of the effect of the sector of employment according to the mode and mean criteria are similar in the entire sample and for males.

As additional robustness checks we introduce other sets of control variables to the baseline specification of the biprobit model. The results for the estimate of the coefficient of the sector of employment and of the correlation of the error terms, ρ, are reported in Table 6.[10] First we add the education of parents as an aim for controlling for the income and wealth of the family of origin. Wealth or family income can be considered to be proxies for the entry to social networks that facilitate information about jobs opportunities and access to employers and might enhance the probability of obtaining a good match between the education acquired and the one required to perform a job. In fact, 70% of our sample responds that by asking help to relatives, friends or colleagues they found their current job. Given the importance of these informal channels through which job search takes place, particularly in developing countries, over-education is less likely to happen to those individuals who have access to social networks and face less information constraints. Once we introduce the parental education variables, all of them display a negative and some are statistically significant, the effect of the formal sector is reduced especially for men, however the coefficient continue to be highly significant.

[9] This discrepancy between the two measurements is usually found in other studies (eg. Kiker et al. 1997; Quinn & Rubb, 2006). It is worth noticing that the correlation between these two methods is usually low, in our particular case we found a correlation of 0.53.

[10] The full set of results is available upon request.

We also add cohort's effects to account for differences across individuals with similar levels of education but exposed to different context. The cohort's variables are not statistical significant and the effect of the formal sector is almost unchanged.

Since there is not a general consensus about how to define and measure informality as a robustness checks we also use alternative measures of informality to evaluate the sensitivity of our results. If we define formal workers as those who make pension contribution or those who make health insurance contribution then our results does not change too much. With pension criterion the effect of formal sector is lower (3pp approximately) and with health criterion the effect is slightly higher. However, if we define formal workers as those who work in establishments of more than 10 workers we find that our results change significantly. In fact, with the size criterion the effect of formal sector is positive. It is important to indicate that once the size criterion is used the controls for the establishment size are eliminated because of collinearity. Therefore, the formal sector variable using the size criterion is equivalent to the controls for the establishment size. Moreover just taking into consideration the size of the establishment to determine if a worker is formal or informal may not be suitable, since it could be the case that workers in small firms are covered by social security system while other employees working in large firms may not.

As a last check we restrict the sample to those with more than 6 years of education (primary complete or more) and those with more than 9 years of education (basic secondary or more). In doing this we obtain that the effect of the formal sector is higher in all cases and statistical significant for the case of women with more than 9 years of education.

7. Conclusions

This study has attempted to extend the over-education literature for developing countries by examine the connection between labor market segmentation, a modern protected formal sector and a traditional-unprotected-informal sector, and over-education in Colombia. So far, studies about informality in developing countries have focused primarily on the size of the informal sector and on the effects of labor market rigidities on employment, wages and its distribution, and on the probability to become informal. However, no attention has been paid to the effects of a large informal sector

in the way workers match their education with the one required to perform their job. This works gives some new evidence into this respect.

In a first step we have shown theoretically that a segmented labor market in two sectors, a modern protected formal sector and a traditional-unprotected-informal sector, can lead to over-education in a developing economy with moderate levels of education. This can explain why skill mismatching is observed in developing countries, such as Colombia, where the average level of educational attainment of the population is still below that in developed countries.

Next, using micro data from Colombia, we estimated two types of empirical models in order to test the relationship between over-education and informality: a simple univariate probit model for the probability of being over-educated that includes as an argument the sector in which the individual is employed, and a bivariate probit model with an endogenous regressor that takes into account that the assignment of workers in each of the sectors is not random and some unobservable factors - ability among others - that influence the probability to choose a particular sector of employment could also affect the probability of being overeducated. The results from the univariate probit estimation indicate that, apparently, formal workers are found less likely to be overeducated rather than having adequate compare to informal workers. However, we have also proved that the assignment of workers in formal or informal sector is not random and that some unobservable characteristics that influence the probability to choose a particular sector of employment also affect the probability of being overeducated, mainly for men. In such scenario, results from the estimation of a standard probit model do not provide consistent results due to sample selection.

The results obtained from the bivariate probit model for the probability of over-education, once potential endogeneity of sector choice and over-education is taken into account, show that formal workers are less likely to be over-educated and that non-observable characteristics that exert a positive effect on the probability of being employed in the formal sector have a positive impact on the probability of being over-educated, for men only. This could be interpreted as evidence that for formal male workers, over-education is caused at least in part by the desire to become part of the formal sector (better employment opportunities, social system protection, etc.). A

worker with considerable formal education possibly will enter a job for which less education is required, because that job is protected, for example, by the minimum wage, while the occupation for which the worker is best suited pays less than the minimum in the informal sector. Nevertheless, given that ability is among the non-observable characteristic this result seems somewhat contradictory. One will expect that workers with more ability, and hence with higher skills, end up working in the formal sector, and that this ability prevents them from being over-educated. Our results imply, therefore, that the negative effect of ability on over-education is more than offset by the positive effect of the desire to become part of the formal sector, among other factors. It is in our agenda, for future research, to explore more in detail the unobservable factors that affect the probability of being over-educated and become part of the formal sector.

Although we are aware that our results have some shortcomings, since for instance it could be argued that better and more suitable instruments can be used, we believe that they are conclusive in terms of correlation and that as a first step for understanding the effect of labor market segmentation on the probability of being over-educated is of importance. According to our results it seems that being a formal worker in addition to the benefits related with social security and probably earning higher wages it also ensures them a better use of their skills in their work. As far as we know there is no study that has shown evidence on this aspect in the past.

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*(2), 259-278.

Alpin, C., Shackleton, J. R., & Walsh, S. (1998). Over- and undereducation in the UK graduate labour market. *Studies in Higher Education, 23*(1), 17-34.

Amaral, P. S., & Quintin, E. (2006). A competitive model of the informal sector. *Journal of Monetary Economics, 53*(7), 1541-1553.

Attanasio, O., Goldberg, P. K., & Pavcnik, N. (2004). Trade reforms and wage inequality in Colombia. *Journal of Development Economics, 74*(2), 331-366.

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*(3), 221-229.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York: Columbia University Press.
Berry, A., & Sabot, R. H. (1978). Labour market performance in developing countries: A survey. *World Development, 6*, 1199-1242.
Buchel, F., & van Ham, M. (2003). Overeducation, regional labor markets, and spatial flexibility. *Journal of Urban Economics, 53*(3), 482-493.
Castillo, M. (2007). Desajuste educativo por regiones en Colombia: ¿competencia por salarios o por puestos de trabajo? In: Cuadernos de Economía, v. XXVI, n. 46, (pp. 107-145). Bogotá.
Charlot, O., & Decreuse, B. (2005). Self-selection in education with matching frictions. *Labour Economics, 12*(2), 251-267.
Chevalier, A. (2003). Measuring Over-education. *Economica, 70*(279), 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. (1972). *Internal Labor Markets and Manpower Analysis*, Lexington, Mass.
Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review, 19*(2), 179-198.
Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review, 1*(1), 75-86.
Duryea, S., Galiani, S., Nopo, H., & Piras, C. C. (2007). The Educational Gender Gap in Latin America and the Caribbean. In: IDB Working Paper No. 502.
Evans, W. N., & Schwab, R. M. (1995). FinishingHighSchoolandStartingCollege: Do Catholic Schools Make a Difference *Quarterly Journal of Economics, 110* 947-974.
Flórez, C. E. (2002). The Function of the Urban Informal Sector in Employment: Evidence from Colombia 1984-2000. In: Documento CEDE No. 2002-04, Universidad de Los Andes. Bogotá D.C.
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*(1), 75-105.
Gong, X., & Van Soest, A. (2001). Wage differentials and mobility in the urban labor market: A panel data analysis for Mexico. In: IZA, Bonn Discussion Paper No. 329.
Green, F., McIntosh, S., & Vignoles, A. (2002). The Utilization of Education and Skills: Evidence from Britain. *The Manchester School, 70*(6), 792-811.
Groot, W., & van den Brink, H. (2000). Skill mismatches in the Dutch labor market. *International Journal of Manpower 21*(8), 584–595.
Hartog, J., & Oosterbeek, H. (1988). Education, allocation and earnings in the Netherlands: 0verschooling? *Economics of Education Review, 7*(2), 185-194.
Kiker, B. F., Santos, M. C., & de Oliveira, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review, 16*(2), 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*.
Magnac, T. (1991). Segmented or competitive labor markets. . *Econometrica, 59* 165–187.
Maloney, W. F. (2004). Informality Revisited. *World Development, 32*(7), 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,* (pp. 109-130). University of Chicago Press and National Bureau of Economic Research.

Mavromaras, K., & McGuinness, S. (2010). Overskilling Dynamics and Education Pathways*. *Economics of Education Review*, doi:10.1016/j.econedurev.2012.1002.1006.

McGoldrick, K., & Robst, J. (1996). Gender differences in overeducation: a test of the theory of differential overqualification. *American Economic Review* (AEA Conference proceedings).

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*(3), 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*(6), 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*(2), 199-206.

Mondragón-Veléz, C., Peña, X., & Wills, D. (2010). Labor Market Rigidities and Informality in Colombia. In: 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, 2 SENA*.

Núñez, J. (2002). Empleo informal y evasión fiscal en Colombia. In *Archivos de Economía, 210, DNP*.

Pradhan, M., & van Soest, A. (1995). Formal and informal sector employment in urban areas of Bolivia. *Labour Economics, 2*(3), 275-297.

Pratap, S., & Quintin, E. (2006). Are labor markets segmented in developing countries? A semiparametric approach. *European Economic Review, 50*(7), 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*(2), 147-156.

Ramos, R., & Sanroma, E. (2011). Overeducation and Local Labour Markets in Spain. In *IZA Discussion Paper 6028*.

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* (2), 101–122.

Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy, 98*(1), 169–192.

Tansel, A. (1999). Formal versus informal sector choice of wage earners and their wages in Turkey. In: Economic Research Forum Working Paper No. 9927.

Tenjo, J. (1990). Opportunities, Aspirations, and Urban Unemployment of Youth: The Case of Colombia. *Economic Development and Cultural Change, 38* (4), 733-761.

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

Tinbergen, J. (1956). On the theory of income distribution. *Weltwirtchaftliches Archiv, 77*, 156–175.

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*(2), 209-228.

Verdugo, R., & Verdugo, N. (1989). The impact of surplus schooling on earnings. *Journal of Human Resources, 24* (4), 629–643.

Verhaest, D., & Omey, E. (2010). The determinants of overeducation: different measures, different outcomes? *International Journal of Manpower, 31*(6), 608 - 625.

Wooldridge, J. M. (2002). *Econometric Analysis of cross section and panel data*, Cambridge, MA: MIT Press.

Table 1. Descriptive statistics for the main variables in the analysis

	Total Sample			Men			Women		
Variable	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal
Over-education	0.15	0.12	0.17	0.16	0.12	0.18	0.15	0.11	0.17
Informal sector	0.33	-	-	0.35	-	-	0.31	-	-
Age (years)	33.93	32.38	34.69	34.09	32.19	35.11	33.71	32.67	34.17
Experience (years)	17.97	18.23	17.85	18.77	18.64	18.84	16.91	17.60	16.60
Tenure (months)	48.56	27.51	59.00	48.69	27.92	59.84	48.39	26.90	57.94
Women	0.43	0.40	0.44	-	-	-	-	-	-
Married	0.53	0.48	0.55	0.61	0.54	0.65	0.41	0.39	0.42
Household Head	0.40	0.37	0.41	0.54	0.47	0.58	0.21	0.23	0.21
Educational Attainment									
Basic Primary or below	0.13	0.23	0.08	0.17	0.28	0.11	0.09	0.16	0.05
Basic secondary	0.14	0.22	0.10	0.17	0.25	0.13	0.10	0.17	0.07
Secondary	0.36	0.37	0.36	0.38	0.34	0.39	0.35	0.41	0.32
Higher education or more	0.36	0.18	0.45	0.28	0.13	0.37	0.46	0.26	0.56
Occupation									
Unskilled manufacture and agricultural	0.33	0.42	0.28	0.48	0.60	0.41	0.13	0.15	0.12
Professionals and Technicians 1	0.07	0.02	0.09	0.07	0.02	0.09	0.06	0.02	0.09
Professionals and Technicians 2	0.05	0.04	0.05	0.03	0.03	0.04	0.07	0.06	0.07
Managers and Public Officials	0.03	0.02	0.04	0.03	0.01	0.04	0.04	0.03	0.04
Administrative Staff	0.20	0.14	0.24	0.14	0.10	0.16	0.28	0.19	0.33
Merchant and Vendor	0.17	0.18	0.16	0.12	0.13	0.12	0.22	0.26	0.21
Service Worker	0.16	0.18	0.15	0.13	0.10	0.14	0.20	0.29	0.15
Firm size									
Micro (1 -10 workers)	0.32	0.68	0.15	0.33	0.68	0.14	0.31	0.67	0.16
Small (11 - 50 workers)	0.22	0.18	0.24	0.22	0.18	0.24	0.22	0.18	0.23
Medium (51 - 100 workers)	0.07	0.03	0.09	0.07	0.04	0.09	0.06	0.03	0.08
Large (101 workers or more)	0.39	0.11	0.53	0.38	0.10	0.53	0.40	0.13	0.53
Sector									
Agricultural, mining, electricity, gas and water	0.03	0.01	0.04	0.04	0.02	0.05	0.02	0.01	0.02
Industry	0.24	0.21	0.25	0.26	0.23	0.28	0.20	0.19	0.20
Construction	0.08	0.14	0.04	0.12	0.23	0.07	0.01	0.01	0.01
Sales, Hotels and Restaurants	0.28	0.36	0.24	0.26	0.31	0.24	0.31	0.44	0.26
Transportation	0.08	0.07	0.09	0.10	0.08	0.10	0.07	0.07	0.07
Financial Intermediation	0.12	0.07	0.14	0.12	0.08	0.14	0.13	0.07	0.15
Social Services	0.17	0.12	0.20	0.10	0.06	0.12	0.26	0.21	0.29
Observations	**15104**	**5006**	**10098**	**8629**	**3013**	**5616**	**6475**	**1993**	**4482**

Note: Figures are in percentages, excepting Age, Experience and Tenure whose units of measurement are indicated in parenthesis.

Table 2. Estimates from the univariate probit over-education model

	Total			Men			Women		
	Coefficient	Marginal Effect	ATE	Coefficient	Marginal Effect	ATE	Coefficient	Marginal Effect	ATE
Formal system	-0.1498**	-0.0250**	-0.0277**	-0.1321*	-0.0186*	-0.0229*	-0.1457*	-0.0272*	-0.0284*
	[0.0384]	[0.0065]	[0.0071]	[0.0518]	[0.0073]	[0.0089]	[0.0579]	[0.0109]	[0.0113]
Schooling years	0.2409**	0.0401**	0.0445**	0.2700**	0.0381**	0.0468**	0.2166**	0.0404**	0.0422**
	[0.0056]	[0.0012]	[0.0009]	[0.0078]	[0.0017]	[0.0012]	[0.0087]	[0.0018]	[0.0015]
Experience (years)	0.0096+	-0.0006+	-0.0002+	-0.001	-0.0015	-0.0014	0.0190*	0.0011*	0.0015*
	[0.0056]	[0.0003]	[0.0004]	[0.0081]	[0.0004]	[0.0005]	[0.0082]	[0.0005]	[0.0006]
Experience2	-0.0004*	-	-	-0.0002	-	-	-0.0004+	-	-
	[0.0001]	-	-	[0.0002]	-	-	[0.0002]	-	-
Tenure (months)	-0.0006	-0.0001	-0.0001	-0.0003	0	0	-0.0009	-0.0002	-0.0002
	[0.0006]	[0.0001]	[0.0001]	[0.0008]	[0.0001]	[0.0001]	[0.0009]	[0.0001]	[0.0001]
Tenure2	0	-	-	0	-	-	0	-	-
	[0.0000]	-	-	[0.0000]	-	-	[0.0000]	-	-
Women	-0.2398**	-0.0400**	-0.0443**	-	-	-	-	-	-
	[0.0419]	[0.0070]	[0.0077]	-	-	-	-	-	-
Married	-0.0232	-0.0039	-0.0043	0.0111	0.0016	0.0019	-0.0505	-0.0094	-0.0098
	[0.0493]	[0.0082]	[0.0091]	[0.0524]	[0.0074]	[0.0091]	[0.0458]	[0.0086]	[0.0089]
Women Married	-0.001	-0.0002	-0.0002	-	-	-	-	-	-
	[0.0661]	[0.0110]	[0.0122]	-	-	-	-	-	-
Household head	-0.049	-0.0082	-0.0091	-0.0283	-0.004	-0.0049	0.0744	0.0139	0.0145
	[0.0493]	[0.0082]	[0.0091]	[0.0525]	[0.0074]	[0.0091]	[0.0532]	[0.0099]	[0.0104]
Women Household head	0.1782*	0.0297*	0.0329*	-	-	-	-	-	-
	[0.0712]	[0.0119]	[0.0132]	-	-	-	-	-	-
Industry	0.0737	0.0123	0.0136	0.1432	0.0202	0.0248	-0.04	-0.0075	-0.0078
	[0.0845]	[0.0141]	[0.0156]	[0.1068]	[0.0150]	[0.0185]	[0.1507]	[0.0281]	[0.0293]
Construction	0.5034**	0.0839**	0.0930**	0.6339**	0.0894**	0.1098**	-0.1767	-0.033	-0.0344
	[0.0983]	[0.0164]	[0.0181]	[0.1207]	[0.0171]	[0.0207]	[0.2187]	[0.0408]	[0.0426]
Sales, Hotels, Restaurants	-0.1029	-0.0171	-0.019	-0.1376	-0.0194	-0.0238	-0.0354	-0.0066	-0.0069
	[0.0848]	[0.0142]	[0.0157]	[0.1082]	[0.0153]	[0.0188]	[0.1490]	[0.0278]	[0.0290]
Transportation	-0.3531**	-0.0588**	-0.0653**	-0.4494**	0.0634**	0.0779**	-0.1926	-0.0359	-0.0375
	[0.0928]	[0.0156]	[0.0171]	[0.1207]	[0.0173]	[0.0209]	[0.1590]	[0.0297]	[0.0310]
Financial Intermediation	-0.4160**	-0.0693**	-0.0769**	-0.5263**	0.0742**	0.0912**	-0.2867+	-0.0535+	-0.0558+
	[0.0895]	[0.0150]	[0.0165]	[0.1167]	[0.0167]	[0.0201]	[0.1533]	[0.0286]	[0.0298]
Social Services	-0.4907**	-0.0818**	-0.0907**	-0.7589**	0.1070**	0.1315**	-0.3402*	-0.0635*	-0.0663*
	[0.0879]	[0.0147]	[0.0161]	[0.1214]	[0.0173]	[0.0207]	[0.1487]	[0.0277]	[0.0289]
Firm Size Small	-0.0271	-0.0045	-0.005	-0.0945	-0.0133	-0.0164	0.0594	0.0111	0.0116
	[0.0428]	[0.0071]	[0.0079]	[0.0584]	[0.0082]	[0.0101]	[0.0640]	[0.0120]	[0.0125]
Firm Size Medium	-0.2150**	-0.0358**	-0.0397**	-0.2305*	-0.0325*	-0.0399*	-0.1645	-0.0307	-0.032
	[0.0662]	[0.0110]	[0.0122]	[0.0896]	[0.0126]	[0.0155]	[0.1012]	[0.0189]	[0.0197]
Firm Size Large	0.0022	0.0004	0.0004	-0.1092+	-0.0154+	-0.0189+	0.1421*	0.0265*	0.0277*
	[0.0414]	[0.0069]	[0.0077]	[0.0573]	[0.0081]	[0.0099]	[0.0609]	[0.0114]	[0.0119]
Metropolitan Area Unemployment Rate	0.0054	0.0009	0.001	0.0049	0.0007	0.0008	0.0082	0.0015	0.0016
	[0.0078]	[0.0013]	[0.0014]	[0.0108]	[0.0015]	[0.0019]	[0.0114]	[0.0021]	[0.0022]
Constant	-3.6861**	-	-	-3.9013**	-	-	-3.8544**	-	-
	[0.1484]	-	-	[0.2000]	-	-	[0.2373]	-	-
Observations	15675			8890			6785		
Wald chi2	2451.21			1569.83			834.53		
Log pseudolikelihood	-5242.92			-2800.57			-2384.24		

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

Table 3. Estimates from the bivariate probit model for the over-education equation.

	Total			Men			Women		
	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE
Formal sector	-0.8406**	-0.1503**	-0.1591**	-1.2063**	-0.2052**	-0.2220**	-0.3354	-0.0624	-0.0651
	[0.2034]	[0.0425]	[0.0411]	[0.1991]	[0.0465]	[0.0412]	[0.2325]	[0.0445]	[0.0455]
Schooling years	0.2508**	0.0449**	0.0475**	0.2748**	0.0468**	0.0506**	0.2214**	0.0412**	0.0429**
	[0.0057]	[0.0023]	[0.0013]	[0.0080]	[0.0030]	[0.0013]	[0.0101]	[0.0024]	[0.0020]
Experience (years)	0.0121*	-0.0003*	0.0002*	0.0083	-0.0009	-0.0005	0.0167*	0.0010*	0.0014*
	[0.0056]	[0.0004]	[0.0004]	[0.0077]	[0.0004]	[0.0006]	[0.0084]	[0.0005]	[0.0006]
Experience2	-0.0004**	-	-	-0.0004+	-	-	-0.0003	-	-
	[0.0001]	-	-	[0.0002]	-	-	[0.0002]	-	-
Tenure (months)	0.0006	0.0001	0.0001	0.0009	0.0001	0.0001	-0.0002	-0.0001	-0.0001
	[0.0007]	[0.0001]	[0.0001]	[0.0008]	[0.0001]	[0.0001]	[0.0011]	[0.0001]	[0.0001]
Tenure2	0.0000	-	-	0.0000	-	-	0.0000	-	-
	[0.0000]	-	-	[0.0000]	-	-	[0.0000]	-	-
Women	-0.2040**	-0.0365**	-0.0386**	-	-	-	-	-	-
	[0.0423]	[0.0074]	[0.0079]	-	-	-	-	-	-
Married	0.027	0.0048	0.0051	0.066	0.0112	0.0122	-0.0567	-0.0106	-0.011
	[0.0528]	[0.0094]	[0.0100]	[0.0540]	[0.0092]	[0.0100]	[0.0464]	[0.0087]	[0.0090]
Women Married	-0.0603	-0.0108	-0.0114	-	-	-	-	-	-
	[0.0685]	[0.0123]	[0.0130]	-	-	-	-	-	-
Household head	-0.068	-0.0122	-0.0129	-0.047	-0.008	-0.0086	0.0845	0.0157	0.0164
	[0.0522]	[0.0093]	[0.0099]	[0.0535]	[0.0091]	[0.0098]	[0.0588]	[0.0109]	[0.0114]
Women Household head	0.2012**	0.0360**	0.0381**	-	-	-	-	-	-
	[0.0770]	[0.0136]	[0.0145]	-	-	-	-	-	-
Industry	0.04	0.0072	0.0076	0.0903	0.0154	0.0166	-0.0483	-0.009	-0.0094
	[0.0851]	[0.0152]	[0.0161]	[0.1045]	[0.0177]	[0.0192]	[0.1516]	[0.0282]	[0.0294]
Construction	0.3974**	0.0711**	0.0752**	0.4008**	0.0682**	0.0738**	-0.0934	-0.0174	-0.0181
	[0.1046]	[0.0178]	[0.0193]	[0.1245]	[0.0200]	[0.0223]	[0.2192]	[0.0408]	[0.0425]
Sales, Hotels, Restaurants	-0.125	-0.0223	-0.0237	-0.1628	-0.0277	-0.03	-0.0275	-0.0051	-0.0053
	[0.0849]	[0.0153]	[0.0161]	[0.1056]	[0.0181]	[0.0194]	[0.1496]	[0.0279]	[0.0290]
Transportation	-0.3718**	-0.0665**	-0.0704**	-0.4597**	-0.0782**	-0.0846**	-0.1948	-0.0363	-0.0378
	[0.0928]	[0.0170]	[0.0176]	[0.1175]	[0.0205]	[0.0216]	[0.1598]	[0.0299]	[0.0310]
Financial Intermediation	-0.4071**	-0.0728**	-0.0770**	-0.4727**	-0.0804**	-0.0870**	-0.2974+	-0.0554+	-0.0577+
	[0.0908]	[0.0162]	[0.0170]	[0.1162]	[0.0197]	[0.0211]	[0.1545]	[0.0287]	[0.0299]
Social Services	-0.5363**	-0.0959**	-0.1015**	-0.8033**	-0.1367**	-0.1479**	-0.3480*	-0.0648*	-0.0675*
	[0.0885]	[0.0166]	[0.0168]	[0.1194]	[0.0217]	[0.0218]	[0.1496]	[0.0280]	[0.0290]
Firm Size Small	0.2298**	0.0411**	0.0435**	0.3128**	0.0532**	0.0576**	0.1433	0.0267	0.0278
	[0.0850]	[0.0167]	[0.0167]	[0.0967]	[0.0193]	[0.0188]	[0.1014]	[0.0193]	[0.0198]
Firm Size Medium	0.1133	0.0203	0.0214	0.2834*	0.0482*	0.0522*	-0.0638	-0.0119	-0.0124
	[0.1127]	[0.0208]	[0.0216]	[0.1271]	[0.0240]	[0.0243]	[0.1410]	[0.0261]	[0.0273]
Firm Size Large	0.3311**	0.0592**	0.0627**	0.4233**	0.0720**	0.0779**	0.2387*	0.0444*	0.0463*
	[0.1020]	[0.0205]	[0.0203]	[0.1127]	[0.0232]	[0.0222]	[0.1186]	[0.0229]	[0.0232]
Metropolitan Area Unemployment Rate	0.0001	0	0	-0.0051	-0.0009	-0.0009	0.0081	0.0015	0.0016
	[0.0080]	[0.0014]	[0.0015]	[0.0105]	[0.0018]	[0.0019]	[0.0119]	[0.0022]	[0.0023]
Constant	-3.5013**	-	-	-3.4764**	-	-	-3.8409**	-	-
	[0.1740]	-	-	[0.2255]	-	-	[0.2478]	-	-
ro	0.3957 **	-	-	0.6076**	-	-	0.1112	-	-
	[0.1355]	-	-	[0.1659]	-	-	[0.1324]	-	-
Observations	15104			8629			6475		
Wald chi2	6556.46			4331.73			2418.11		
Log pseudolikelihood	-11407.29			-6355.20			-4968.14		

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

Table A1. Estimates from the bivariate probit model for the sector equation (formal=1)

	Total		Men		Women	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Schooling years	0.0762**	0.0449**	0.0622**	0.0468**	0.0959**	0.0412**
	[0.0051]	[0.0023]	[0.0062]	[0.0030]	[0.0082]	[0.0024]
Experience (years)	0.0290**	-0.0003**	0.0331**	-0.0009**	0.0254**	0.0010**
	[0.0042]	[0.0004]	[0.0057]	[0.0004]	[0.0064]	[0.0005]
Experience^2	-0.0005**	-	-0.0005**	-	-0.0005**	-
	[0.0001]	-	[0.0001]	-	[0.0002]	-
Tenure (months)	0.0090**	0.0001**	0.0070**	0.0001**	0.0118**	-0.0001**
	[0.0005]	[0.0001]	[0.0007]	[0.0001]	[0.0009]	[0.0001]
Tenure^2	-0.0000**	-	-0.0000**	-	-0.0000**	-
	[0.0000]	-	[0.0000]	-	[0.0000]	-
Women	0.0945*	-0.0365*	-	-	-	-
	[0.0390]	[0.0074]	-	-	-	-
Married	0.1113*	0.0048*	0.0884+	0.0112+	-0.0453	-0.0106
	[0.0482]	[0.0094]	[0.0488]	[0.0092]	[0.0449]	[0.0087]
Women Married	-0.1834**	-0.0108**	-	-	-	-
	[0.0622]	[0.0123]	-	-	-	-
Household head	0.0175	-0.0122	0.004	-0.008	-0.0626	0.0157
	[0.0473]	[0.0093]	[0.0474]	[0.0091]	[0.0534]	[0.0109]
Women Household head	-0.1365*	0.0360*	-	-	-	-
	[0.0668]	[0.0136]	-	-	-	-
Industry	-0.1640+	0.0072+	-0.1413	0.0154	-0.1626	-0.009
	[0.0890]	[0.0152]	[0.1018]	[0.0177]	[0.1720]	[0.0282]
Construction	-0.4391**	0.0711**	-0.4814**	0.0682**	-0.3379	-0.0174
	[0.0956]	[0.0178]	[0.1064]	[0.0200]	[0.2363]	[0.0408]
Sales, Hotels and Restaurants	-0.0905	-0.0223	-0.064	-0.0277	-0.093	-0.0051
	[0.0889]	[0.0153]	[0.1020]	[0.0181]	[0.1706]	[0.0279]
Transportation	-0.1515	-0.0665	-0.1524	-0.0782	-0.1651	-0.0363
	[0.0960]	[0.0170]	[0.1105]	[0.0205]	[0.1828]	[0.0299]
Financial Intermediation	0.1471	-0.0728	0.1414	-0.0804	0.1534	-0.0554
	[0.0960]	[0.0162]	[0.1125]	[0.0197]	[0.1797]	[0.0287]
Social Services	-0.3248**	-0.0959**	-0.3389**	-0.1367**	-0.3142+	-0.0648+
	[0.0922]	[0.0166]	[0.1125]	[0.0217]	[0.1710]	[0.0280]
Firm Size Small	1.0249**	0.0411**	1.0538**	0.0532**	0.9918**	0.0267**
	[0.0319]	[0.0167]	[0.0417]	[0.0193]	[0.0498]	[0.0193]
Firm Size Medium	1.4104**	0.0203**	1.3984**	0.0482**	1.4514**	-0.0119**
	[0.0555]	[0.0208]	[0.0696]	[0.0240]	[0.0925]	[0.0261]
Firm Size Large	1.6818**	0.0592**	1.7397**	0.0720**	1.6061**	0.0444**
	[0.0331]	[0.0205]	[0.0448]	[0.0232]	[0.0505]	[0.0229]
Metropolitan Area Unemployment Rate	-0.0277**	0.0000**	-0.0181*	-0.0009*	-0.0414**	0.0015**
	[0.0069]	[0.0014]	[0.0091]	[0.0018]	[0.0105]	[0.0022]
Mean Yedu	0.0320**	0.0000**	0.0378**	0.0000**	0.0250**	0.0000**
	[0.0047]	[0.0000]	[0.0059]	[0.0000]	[0.0073]	[0.0000]
Num. Chidren 0 - 1 years old	-0.1019*	0.0000*	-0.1148*	0.0000*	-0.0642	0.0000
	[0.0439]	[0.0000]	[0.0527]	[0.0000]	[0.0764]	[0.0000]
Constant	-1.6359**	-	-1.6741**	-	-1.5469**	-
	[0.1369]	-	[0.1716]	-	[0.2354]	-
Observations	15104		8629			6475
Wald chi2	6556.46		4331.73			2418.11
Log pseudolikelihood	-11407.29		-6355.20			-4968.14

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

Table 4. Estimates from the univariate probit over-education model – Mode method

	Total			Men			Women		
	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE
Formal system	-0.4597**	0.1261**	0.1113**	-0.7650**	-0.1915**	-0.1737**	0.6604+	0.1997+	0.1700+
	[0.1142]	[0.0325]	[0.0278]	[0.1268]	[0.0347]	[0.0294]	[0.3518]	[0.1075]	[0.0911]
Schooling years	0.2805**	0.0770**	0.0679**	0.2754**	0.0690**	0.0625**	0.2739**	0.0828**	0.0705**
	[0.0060]	[0.0019]	[0.0010]	[0.0079]	[0.0023]	[0.0012]	[0.0189]	[0.0053]	[0.0043]
Experience (years)	-0.0193**	0.0018**	0.0021**	-0.0104	-0.0009	-0.0011	-0.0281**	0.0033**	0.0035**
	[0.0049]	[0.0005]	[0.0005]	[0.0066]	[0.0006]	[0.0006]	[0.0074]	[0.0008]	[0.0007]
Experience2	0.0004**	-	-	0.0002	-	-	0.0005**	**	**
	[0.0001]	-	-	[0.0002]	-	-	[0.0002]		
Tenure (months)	0.0008	0.0002	0.0002	0.0001	0	0	-0.0007	-0.0001	-0.0001
	[0.0006]	[0.0001]	[0.0001]	[0.0007]	[0.0001]	[0.0001]	[0.0012]	[0.0003]	[0.0002]
Tenure2	0	-	-	0	-	-	0	-	-
	[0.0000]	-	-	[0.0000]	-	-	[0.0000]	-	-
Women	0.0867*	0.0238*	0.0210*	-	-	-	-	-	-
	[0.0381]	[0.0104]	[0.0092]	-	-	-	-	-	-
Married	-0.0723	-0.0198	-0.0175	-0.0743	-0.0186	-0.0169	-0.0718+	-0.0217+	-0.0185+
	[0.0503]	[0.0138]	[0.0122]	[0.0500]	[0.0126]	[0.0113]	[0.0405]	[0.0123]	[0.0104]
Women Married	0.0076	0.0021	0.0018				-	-	-
	[0.0625]	[0.0172]	[0.0151]				-	-	-
Household head	0.0243	0.0067	0.0059	0.0366	0.0092	0.0083	0.0164	0.005	0.0042
	[0.0498]	[0.0137]	[0.0121]	[0.0495]	[0.0124]	[0.0112]	[0.0522]	[0.0158]	[0.0134]
Women Household head	-0.0361	-0.0099	-0.0087	-	-	-	-	-	-
	[0.0690]	[0.0189]	[0.0167]	-	-	-	-	-	-
Industry	-0.1575+	-0.0432+	-0.0381+	-0.1238	-0.031	-0.0281	-0.1467	-0.0444	-0.0378
	[0.0847]	[0.0233]	[0.0205]	[0.1038]	[0.0261]	[0.0236]	[0.1393]	[0.0421]	[0.0358]
Construction	1.1055**	0.3033**	0.2676**	1.0586**	0.2651**	0.2403**	0.0313	0.0095	0.0081
	[0.1060]	[0.0285]	[0.0249]	[0.1244]	[0.0302]	[0.0271]	[0.2238]	[0.0677]	[0.0576]
Sales, Hotels and Restaurants	-0.1468+	-0.0403+	-0.0355+	-0.2184*	-0.0547*	-0.0496*	0.021	0.0064	0.0054
	[0.0842]	[0.0232]	[0.0204]	[0.1045]	[0.0263]	[0.0238]	[0.1352]	[0.0409]	[0.0348]
Transportation	-0.1972*	-0.0541*	-0.0477*	-0.3053**	-0.0764**	-0.0693**	0.0213	0.0064	0.0055
	[0.0893]	[0.0246]	[0.0216]	[0.1106]	[0.0279]	[0.0252]	[0.1457]	[0.0441]	[0.0375]
Financial Intermediation	-0.1096	-0.0301	-0.0265	-0.1823+	-0.0457+	-0.0414+	-0.0446	-0.0135	-0.0115
	[0.0873]	[0.0240]	[0.0211]	[0.1099]	[0.0275]	[0.0250]	[0.1398]	[0.0423]	[0.0360]
Social Services	-0.3559**	0.0977**	0.0862**	-0.3546**	-0.0888**	-0.0805**	-0.2646+	-0.0800+	-0.0681+
	[0.0863]	[0.0238]	[0.0209]	[0.1115]	[0.0280]	[0.0253]	[0.1398]	[0.0422]	[0.0359]
Firm Size Small	0.1327*	0.0364*	0.0321*	0.1881**	0.0471**	0.0427**	-0.1733	-0.0524	-0.0446
	[0.0564]	[0.0157]	[0.0137]	[0.0719]	[0.0185]	[0.0164]	[0.1394]	[0.0424]	[0.0360]
Firm Size Medium	0.1248	0.0342	0.0302	0.1894+	0.0474+	0.0430+	-0.2317	-0.0701	-0.0596
	[0.0772]	[0.0214]	[0.0187]	[0.0976]	[0.0249]	[0.0223]	[0.1857]	[0.0565]	[0.0480]
Firm Size Large	0.2775**	0.0761**	0.0672**	0.3542**	0.0887**	0.0804**	-0.1312	-0.0397	-0.0338
	[0.0653]	[0.0185]	[0.0159]	[0.0802]	[0.0211]	[0.0184]	[0.1864]	[0.0566]	[0.0481]
Metropolitan Area Unemployment Rate	0.005	0.0014	0.0012	-0.0143	-0.0036	-0.0032	0.0311**	0.0094**	0.0080**
	[0.0070]	[0.0019]	[0.0017]	[0.0096]	[0.0024]	[0.0022]	[0.0105]	[0.0032]	[0.0027]
Constant	-3.5615**	-	-	-3.1352**	-	-	-4.2167**	-	-
	[0.1490]	-	-	[0.1937]	-	-	[0.2358]	-	-
athrho	0.2126**	-	-	0.3394**	-	-	-0.3852	-	-
	[0.0663]	-	-	[0.0760]	-	-	[0.2398]	-	-
Observations	15104			8629			6475		
Wald chi2	7009.23			4137.57			3132.86		
Log pseudolikelihood	-12862.27			-7114.49			-5646.93		

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

Table 5. Estimates from the bivariate probit model for the over-education equation – Mode method

	Total			Men			Women		
	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE	Probit Coefficient	Marginal Effect	ATE
Formal sector	-0.4597**	-0.1261**	-0.1113**	-0.7650**	-0.1915**	-0.1737**	0.6604+	0.1997+	0.1700+
	[0.1142]	[0.0325]	[0.0278]	[0.1268]	[0.0347]	[0.0294]	[0.3518]	[0.1075]	[0.0911]
Schooling years	0.2805**	0.0770**	0.0679**	0.2754**	0.0690**	0.0625**	0.2739**	0.0828**	0.0705**
	[0.0060]	[0.0019]	[0.0010]	[0.0079]	[0.0023]	[0.0012]	[0.0189]	[0.0053]	[0.0043]
Experience	-0.0193**	-0.0018**	-0.0021**	-0.0104	-0.0009	-0.0011	-0.0281**	0.0033**	0.0035**
	[0.0049]	[0.0005]	[0.0005]	[0.0066]	[0.0006]	[0.0006]	[0.0074]	[0.0008]	[0.0007]
Experience2	0.0004**	-	-	0.0002	-	-	0.0005**	**	**
	[0.0001]	-	-	[0.0002]	-	-	[0.0002]		
Tenure (months)	0.0008	0.0002	0.0002	0.0001	0	0	-0.0007	-0.0001	-0.0001
	[0.0006]	[0.0001]	[0.0001]	[0.0007]	[0.0001]	[0.0001]	[0.0012]	[0.0003]	[0.0002]
Tenure2	0	-	-	0	-	-	0	-	-
	[0.0000]	-	-	[0.0000]	-	-	[0.0000]	-	-
Women	0.0867*	0.0238*	0.0210*	-	-	-	-	-	-
	[0.0381]	[0.0104]	[0.0092]	-	-	-	-	-	-
Married	-0.0723	-0.0198	-0.0175	-0.0743	-0.0186	-0.0169	-0.0718+	-0.0217+	-0.0185+
	[0.0503]	[0.0138]	[0.0122]	[0.0500]	[0.0126]	[0.0113]	[0.0405]	[0.0123]	[0.0104]
Women Married	0.0076	0.0021	0.0018				-	-	-
	[0.0625]	[0.0172]	[0.0151]				-	-	-
Household head	0.0243	0.0067	0.0059	0.0366	0.0092	0.0083	0.0164	0.005	0.0042
	[0.0498]	[0.0137]	[0.0121]	[0.0495]	[0.0124]	[0.0112]	[0.0522]	[0.0158]	[0.0134]
Women Household head	-0.0361	-0.0099	-0.0087	-	-	-	-	-	-
	[0.0690]	[0.0189]	[0.0167]	-	-	-	-	-	-
Industry	-0.1575+	-0.0432+	-0.0381+	-0.1238	-0.031	-0.0281	-0.1467	-0.0444	-0.0378
	[0.0847]	[0.0233]	[0.0205]	[0.1038]	[0.0261]	[0.0236]	[0.1393]	[0.0421]	[0.0358]
Construction	1.1055**	0.3033**	0.2676**	1.0586**	0.2651**	0.2403**	0.0313	0.0095	0.0081
	[0.1060]	[0.0285]	[0.0249]	[0.1244]	[0.0302]	[0.0271]	[0.2238]	[0.0677]	[0.0576]
Sales, Hotels and Restaurants	-0.1468+	-0.0403+	-0.0355+	-0.2184*	-0.0547*	-0.0496*	0.021	0.0064	0.0054
	[0.0842]	[0.0232]	[0.0204]	[0.1045]	[0.0263]	[0.0238]	[0.1352]	[0.0409]	[0.0348]
Transportation	-0.1972*	-0.0541*	-0.0477*	-0.3053**	-0.0764**	-0.0693**	0.0213	0.0064	0.0055
	[0.0893]	[0.0246]	[0.0216]	[0.1106]	[0.0279]	[0.0252]	[0.1457]	[0.0441]	[0.0375]
Financial Intermediation	-0.1096	-0.0301	-0.0265	-0.1823+	-0.0457+	-0.0414+	-0.0446	-0.0135	-0.0115
	[0.0873]	[0.0240]	[0.0211]	[0.1099]	[0.0275]	[0.0250]	[0.1398]	[0.0423]	[0.0360]
Social Services	-0.3559**	-0.0977**	-0.0862**	-0.3546**	-0.0888**	-0.0805**	-0.2646+	-0.0800+	-0.0681+
	[0.0863]	[0.0238]	[0.0209]	[0.1115]	[0.0280]	[0.0253]	[0.1398]	[0.0422]	[0.0359]
Firm Size Small	0.1327*	0.0364*	0.0321*	0.1881**	0.0471**	0.0427**	-0.1733	-0.0524	-0.0446
	[0.0564]	[0.0157]	[0.0137]	[0.0719]	[0.0185]	[0.0164]	[0.1394]	[0.0424]	[0.0360]
Firm Size Medium	0.1248	0.0342	0.0302	0.1894+	0.0474+	0.0430+	-0.2317	-0.0701	-0.0596
	[0.0772]	[0.0214]	[0.0187]	[0.0976]	[0.0249]	[0.0223]	[0.1857]	[0.0565]	[0.0480]
Firm Size Large	0.2775**	0.0761**	0.0672**	0.3542**	0.0887**	0.0804**	-0.1312	-0.0397	-0.0338
	[0.0653]	[0.0185]	[0.0159]	[0.0802]	[0.0211]	[0.0184]	[0.1864]	[0.0566]	[0.0481]
Metropolitan Area Unemployment Rate	0.005	0.0014	0.0012	-0.0143	-0.0036	-0.0032	0.0311**	0.0094**	0.0080**
	[0.0070]	[0.0019]	[0.0017]	[0.0096]	[0.0024]	[0.0022]	[0.0105]	[0.0032]	[0.0027]
Constant	-3.5615**	-	-	-3.1352**	-	-	-4.2167**	-	-
	[0.1490]	-	-	[0.1937]	-	-	[0.2358]	-	-
ρ	0.2126**	-	-	0.3394**	-	-	-0.3852	-	-
	[0.0663]	-	-	[0.0760]	-	-	[0.2398]	-	-
Observations	15104			8629			6475		
Wald chi2	7009.23			4137.57			3132.86		
Log pseudolikelihood	-12862.27			-7114.49			-5646.93		

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

Table 6: Additional Robustness Checks

Total Sample

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Formal sector	-0.1503**	-0.1050**	-0.1028**	-0.1261**	-0.1133**	-0.1658**	0.2090**	-0.2030**	-0.2700**
	[0.0425]	[0.0375]	[0.0381]	[0.0325]	[0.0416]	[0.0474]	[0.0353]	[0.0537]	[0.0475]
ρ	0.3957 **	0.3330*	0.3198*	0.2126**	0.2935*	0.4791**	-0.8916**	0.4109**	0.4768**
	[0.1355]	[0.1407]	[0.1430]	[0.0663]	[0.1379]	[0.1583]	[0.1512]	[0.1313]	[0.1052]
Individual's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Job's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parents' education		Yes	Yes						
Cohorts			Yes						
Mode				Yes					
Pension					Yes				
Health						Yes			
Size							Yes		
Years education > 6								Yes	
Years education > 9									Yes
Observations	15675	9333	9333	15104	15104	15104	15104	12679	10975

Men

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Formal sector	-0.2052**	-0.1265**	-0.1161**	-0.1915**	-0.1721**	-0.2777**	0.2471**	-0.3049**	-0.3666**
	[0.0465]	[0.0546]	[0.0558]	[0.0347]	[0.0452]	[0.0319]	[0.0218]	[0.0631]	[0.0585]
ρ	0.6076**	0.5283*	0.4913+	0.3394**	0.5812**	1.0641**	-1.3777**	0.6830**	0.6691**
	[0.1659]	[0.2380]	[0.2565]	[0.0760]	[0.1649]	[0.1405]	[0.1420]	[0.1670]	[0.1371]
Individual's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Job's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parents' education		Yes	Yes						
Cohorts			Yes						
Mode				Yes					
Pension					Yes				
Health						Yes			
Size							Yes		
Years education > 6								Yes	
Years education > 9									Yes
Observations	8629	5642	5642	8629	8629	8629	8629	6882	5695

Women

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Formal sector	-0.0624	-0.071	-0.0767	0.1997+	-0.0247	0.011	0.1636**	-0.0797	-0.1149*
	[0.0445]	[0.0526]	[0.0554]	[0.1075]	[0.0405]	[0.0383]	[0.0518]	[0.0538]	[0.0584]
ρ	0.1112	0.1868	0.1977	-0.3852	-0.0102	-0.1171	-0.5351**	0.121	0.1749
	[0.1324]	[0.1710]	[0.1747]	[0.2398]	[0.1248]	[0.1207]	[0.1721]	[0.1319]	[0.1286]
Individual's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Job's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parents' education		Yes	Yes						
Cohorts			Yes						
Mode				Yes					
Pension					Yes				
Health						Yes			
Size							Yes		
Years education > 6								Yes	
Years education > 9									Yes
Observations	6475	3691	3691	6475	6475	6475	6475	5797	5280

Notes: Marginal effects of a biprobit model are presented with different covariates, column [1] is the baseline model presented in Table 3. Standard errors in []. + p<0.1, * p<0.05, ** p<0.01