Three Empirical Essays on Education and Informality in the Labor Market of a Developing Country: The Colombian Case

Paula Herrera-Idárraga

Three Empirical Essays on Education and Informality in the Labor Market of a Developing Country: The Colombian Case

Paula Herrera-Idárraga

Supervisors

Enrique López-Bazo
Elisabet Motellón

Thesis submitted for the degree of Doctor of Philosophy
University of Barcelona 2014

Acknowledgments

I would like to express my most sincere gratitude to my two thesis supervisors, Dr. Elisabet Motellón and Dr. Enrique López-Bazo. Eli always encouraged me to believe in myself and in my ideas (not an easy task, I have to say). She was a constant guide for developing these ideas in the most rigorous way. She also made possible that Dr. López-Bazo accepted to be my supervisor. I quickly came to understand why having Quique as my supervisor was fundamental for developing my thesis. This thesis would never have been possible without his gigantic patience, and his constant guidance for assessing every methodological aspect. I have to say, that having not one but two demanding and critical supervisors has been an interesting and very constructive experience. There will always remain etched on my mind each of the meetings in which Eli and Quique were present. Not only because in each of these meetings was where I learned the most, but also where I had the best time doing so.

I am greatly indebted to AQR research group and the Department of Econometrics, Statistics and Spanish Economy at the University of Barcelona for hosting me and for the support during these years. I expressly want to thank Dr. Manuel Artis who helped me to get funding for my studies. I would not have had the possibility to be exclusively devoted to my thesis without the funding provided by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) and the support from the Pontificia Universidad Javeriana.

I also want to thank all my colleagues who suffered the same penalties and sorrows I endured during this time, my partners in the master and doctorate program. Thanks David for sharing with me the office, for all the conversations we had about our research topics and for your fruitful comments. My best memories of meals and breaks with my two inseparable companions and friends: Ana and Arelly, thanks for being all ears on my most critical moments. And of course I will never forget my beloved *trio of musketeers*: Nati, Paola and Dario at UAB.

Special thanks to some of the people I lived with during my stay in the Barcelona: Franceline, Laura, Valentina and Oscar. Thanks to you all I could escape from my thoughts and the thesis when I needed to and had a place just like home. Laura, thank you for making me feel like family. I also want to thank some people we met again, Adri and Anna, and some other people that appeared in the most unexpected places and that will be remembered forever, Sonia and Ivan.

I am also thankful to the people that despite the distance supported me all this time. My two best friends, Nico and Lole. I know you share with me all the joy of having completed this thesis.

Finally I would like to thank my family. My sister Carolina who made it possible for me to perform my studies abroad without worrying about all the responsibilities I left behind. Thanks Caro for your constant support. Special thanks to my father and my mother. They both have been my support for pursuing the doctorate and finishing the thesis. My father who always had confidence that I would end this thesis. So much that since the day I started studying, every day we talked he asked me if I had already finished it. My mother who has given me all her unconditional love and affection. *Gracias papá y mamá.*

Abstract

This dissertation consists of three essays with a marked empirical orientation. The first two essays provide empirical evidence concerning the relationship between informality and education-occupation mismatches in a developing country. While the third chapter analyze regional wage inequalities in a developing country and the role of education and informality. The three essays of the dissertation are entitled: "Informality and Overeducation in the Labor Market of a Developing Country", "Double Penalty in Returns to Education: Informality and Educational Mismatch in the Colombian Labor Market" and "Wage Gaps Across Colombian Regions: The Role of Education and Informality"

Informality and Overeducation in the Labor Market of a Developing Country (co-authored with Enrique López-Bazo and Elisabeth Motellón)

This chapter explores the connection between labor market segmentation in two sectors, a modern protected formal sector and a traditional- unprotected-informal sector, and overeducation in a developing country. Informality is thought to have negative consequences, primarily through poorer working conditions, lack of social security, as well as low levels of productivity throughout the economy. This chapter considers an aspect that has not been previously addressed, namely the fact that informality might also affect the way workers match their actual education with that required performing their job. Using micro-data from Colombia the relationship between overeducation and informality is tested. Empirical results suggest that, once the endogeneity of employment choice has been accounted for, formal male workers are less likely to be overeducated. Interestingly, the propensity of being overeducated among women does not seem to be closely related to the sector choice.

Double Penalty in Returns to Education: Informality and Educational Mismatch in the Colombian Labor Market (co-authored with Enrique López-Bazo and Elisabeth Motellón)

This chapter examines the returns to education taking into consideration the existence of educational mismatches in the formal and

informal employment of a developing country. Results show that the returns of surplus, required and deficit years of schooling are different in the two sectors. Moreover, they suggest that these returns vary along the wage distribution, and that the pattern of variation differs for formal and informal workers. In particular, informal workers face not only lower returns to their education, but suffer a second penalty associated with educational mismatches that puts them at a greater disadvantage compare to their formal counterparts.

Wage Gaps Across Colombian Regions: The Role of Education and Informality (co-authored with Enrique López-Bazo and Elisabeth Motellón)

This chapter analyzes the role of education and informality on regional wage differentials. The hypothesis that is put under examination is that apart from the difference in the endowments of human capital across regions, regional heterogeneity in the incidence of informality may be another important source of regional wage inequality. The results for Colombian regions confirm marked differences in wage distributions between regions and that they differ in the endowment of human capital and more importantly in the incidence of informality. Regional heterogeneity in returns to education is especially intense in the upper part of the wage distribution. While heterogeneity in the informal pay penalty throughout the territory is more relevant in the lower part of the wage distribution.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Increased investment in education is often promoted as a key development strategy, aimed generally at boosting economic growth and poverty reduction. This is due in part to the fact that the opportunities a person has for escaping poverty are determined by how easily it is to obtain higher levels of education, since the latter are closely linked to higher earnings (Mincer 1970). Human capital theories argue that it is possible to increase productivity and hence income through greater investment in education. In fact, education is closely related to the increase of labor force participation, a higher chance to find a job and get higher wages. Furthermore, in addition to generating private returns in the form of higher wages, education has also other non-monetary returns for the society as a whole associated with health, fertility and crime.

Following this idea that more education is a crucial factor for economic growth and development, Latin America countries have done great efforts at increasing the education levels of their population. In the last two decades these countries have experienced substantial changes in the educational attainment levels of its labor force. According to World Bank data, the percentage of students enrolled in tertiary education has roughly doubled in the last 10 years, from 22.8% in 2000 to 42.3% in 2011. Definitely educational improvement is good news for the region. However, while the increase in educational attainment might be worth just because of its non-pecuniary benefits, it might also be desirable that these additional investments in human capital are productively used in the economy. This last point is one of the centerpieces of this dissertation; it will be examined under what context some educational skills might be underutilized in a developing country and the consequences of this misallocation over the returns to education.

We will claim that given the particularities of labor markets in some developing countries, and under certain conditions, returns to additional investment in education may face some constraints and thus all its potential is not obtained. One of this characteristics is the fact that in almost all Latin American and the Caribbean labor markets there is the existence and the persistence of a large informal sector. Actually, half of the employed population of this region worked in informal jobs at the end of the first decade of this century (International Labour Organization [ILO], 2011). Informal employment embraces a variety of heterogeneous activities, such as self-employment entrepreneurs, salaried workers of large and small firms, family and domestic workers. Informal employment generally involves that workers are trapped in unproductive activities, with inferior working conditions, lack of social security and lower earnings.

Alternative definitions and corresponding ways of measuring informality have been proposed in the literature. This lack of consensus largely reflects issues of data availability in each country under study. There are two well-known ways for defining informality. The first is the "productive" definition, where informality is associated to firms that operate at a small-scale, have low-productivity and are frequently family-based (Maloney, 2004). However given that productivity is not easily observable, the "productive" definition has been reduced to the easiest observable characteristic of the firm and correlated with productivity, its size. According to which, self-employed, workers employed in firms of 5 or less employees and unpaid family workers are considered to be informal workers. This definition has been criticized in the literature because it does not take into account the benefits associated with formal employment, such as inclusion in the social security system (Flórez 2000). For instance, it is possible that employees of large firms are not covered by the social security system. Thus, the second definition emphasizes social regulation. According to this definition an informal job is that of a job that does not pay contributions to the social security system. Throughout this dissertation a worker is consider informal employee if he does not contribute to both health and pension systems. This definition is also in line with the definition proposed by the Seventeenth International Conferences of Labor Statisticians (ICLS).[1] More importantly, because data usually comes from household surveys and thus the information relates only to workers and not of the firm, the informal sector term is related to the nature of the job and not of the firm in which the worker is employed (Botelho and Ponczek, 2011).

There exist different reasons that can explain why informal workers are rewarded differently from identical formal workers. According to the dualistic view, based on the Harris and Todaro (1970) model, jobs are rationed in the formal sector due to labor market rigidities of an institutional nature, such as labor unions and minimum wages legislation. As a result some workers are forced to accept informal sector jobs. A seemingly stylized fact, found in past studies about labor market segmentation, is that informal-sector workers, even if equally productive, are subject to lower remuneration than formal-sector workers. In fact, most studies of labor markets in developing countries find that some characteristics are better rewarded in formal jobs (e.g., Pradhan and Van Soest 1995; Tansel 2000; Gong and Van Soest 2002; Botelho and Ponczek 2011).

However, several recent studies postulate that, for both firms and workers, the decision of being formal turns out to be extremely costly, due to

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

the non labor costs associated with health and pension contributions, payroll taxes, commuting subsidies, among others, which significantly increases the attractiveness of informal activities. Maloney (1999), for instance, introduces a standpoint in which workers may find informal-sector employment a desirable alternative, due to inefficiencies in the provision of public services, that is, health and pension, or because their level of human capital does not fulfill the requirements for performing formal jobs. In the last case, a wage penalty for informal-sector employment may be due to sorting, where those with low levels of human capital are also those more likely to work in the informal sector (Tokman, 1982). This type of sorting may result from the fact that firms in the informal sector have limited access to financing and employers choose to substitute physical capital for low-skill labor (see, for example, Amaral and Quintin, 2006).

Recent theories of labor market segmentation try to reconcile the two opposite views regarding the informal sector. Fields (1990 and 2005) postulate that the informal sector is segmented, in turn, into two segments. One segment in which, informal jobs are preferred to certain formal jobs, labeled as the "upper tier". While in the other segment, the "lower tier", corresponds to informal jobs that offer worse working conditions than those offered in similar formal jobs.

All in all, no matter the reason which best describes the existence of the informal sector, there is not a doubt that its presence and persistence in most of Latin America economies must have crucial repercussions on the wage structure, on the performance of labor markets overall and eventually on the incentives for individuals to continue accumulating further education.

On the other hand, in several developed countries, and recently in some developing countries, a feature of concern is the discrepancy between education acquired by workers and skill requirements of jobs, commonly known as educational mismatch. In particular, the phenomenon of overeducation has been extensively studied. An individual worker is said to be overeducated if she has acquired more education than what is required to perform her job. Overeducation is, thus, often taken to imply that resources are not efficiently used, since overeducated workers make lower returns on their investment relative to similarly educated individuals whose jobs match appropriately their level of education (for an extensive review of overeducation in developed countries see McGuinness, 2006 and Leuven and Oosterbeek, 2011). One concern is that the increase in the educational attainment of the labor force in some Latin America countries may have overtaken the growth of jobs for high skilled workers. If the growth of formal jobs, usually related with high skills requirements, has not increased at the same rate as educational attainment of the workforce, then workers with higher education may have displaced low skilled workers from jobs that possess lower skill requirements, usually available in the informal sector. Thus workers skills might exceed the skills required for performing the jobs.

One of the main hypothesis of this dissertation is that overeducation in a developing country is not independent of market segmentation into formal and informal sector. It is possible that labor market segmentation might affect the way workers match their actual education with the one required to perform their job. So even when more highly educated workers tend to be more productive than less skilled counterparts, education may not be the key for higher paying jobs if the labor market is segmented. The hypothesis that will be tested is that, a highly skilled worker who is unable to obtain a high-skill job in the formal sector, may accept a low-skill job in the informal sector for which she is overeducated. Furthermore it will be examined if educational mismatches can explain, at least in part, the wage gap between formal and informal workers. This dissertation provides new relevant evidence on the relationship between overeducation and informality in a developing country.

Up to now studies about informality for developing countries focus primarily on the size of the informal sector, on the likelihood that workers enter/exit the informal sector and on the formal/informal sector wage differentials. However, little attention has been paid to the effects of a large informal sector on the way workers match their education with the one required to perform their job. This dissertation seeks to fill this gap in the literature.

Another feature that is closely examined in this dissertation is regional wage inequalities and the role of education and informality in explaining these differences. Educational expansion in Latin American countries has not been homogenous. For example, Cruces et al. (2011) report considerable differences in the average years of education of the adult population positioned in the top quantile compared to those in the bottom quintile. They reported that in some countries this difference could be as large as 7.5 years. Moreover, Latin American countries display important differences in the distribution of human capital throughout their territories. Educational expansion was usually concentrated in the capital city and in other main cities, so that other cities, usually the peripherals, were left behind in the process. It is possible to find, within a country, regions with the highest percentage of individuals with the highest level of education, and regions with the highest percentage of individuals with the lowest levels of education. Aside from these differences in the dotation of human capital across regions, several studies have found that returns to education differ across territories (Azzoni and Servo, 2002; Romero, 2008; Quiñones and Rodriguez, 2011). Then regional wage inequalities can be generated by these disparities in dotation and returns. Apart from the difference in the endowments across regions, regional heterogeneity in the incidence of informality may be another important source of regional wage inequality. The hypothesis that will be tested is that regional wage inequality may be explained, at least partly, by regional differences in the availability of formal jobs, i.e. good jobs that generate higher wages. Into this regard, this dissertation will give some new empirical evidence concerning

regional wage inequalities in a developing country and the relationship of these inequalities with informality and education.

The Colombian case has been selected because is a good example of a developing country characterized by a high degree of informality in its labor market. And because its features make the Colombian case to be representative in regard to the situation suffered by other developing countries, with similar characteristics, concerning the problem of informality. The country's informal employment is an interesting case to study for several reasons. First, informality today is at the center of economic debate in the country because of the high levels that prevail, around half of the working force has an informal job. As a matter of fact the government confronted with persistently informality and high unemployment rate among youth, had released a new law, "Ley de Formalización y Generación de Empleo" (Law 1429 of 2010), that was aimed at improving the employment situation of young people in two fronts, job creation and the quality of employment. Second, in Colombia there is a high incidence of the minimum wage, i.e. a relatively high proportion of formal sector employees receive a salary similar to the minimum, which points to the existence of important labor market rigidities. Third, informality rates along the Colombian territory are very dissimilar; while some regions present an incidence of informality of around 30% others display an incidence of 70%. Finally, Colombia has also been recognized as a country with vast heterogeneity of its workforce in terms of education along its territory and due to its high levels of regional wage inequality.

Data from the Colombia Household Surveys (CHS) was used. The CHS is a repeated cross-section conducted by the National Statistics Department (DANE). This survey gathers information about employment conditions for a population aged 12 years or more and includes data about income, occupation, industry, and firm size, in addition to the individual's general characteristics of sex, age, marital status and educational attainment. Certain household characteristics, such as the head of the household, the number of children, and the level of education of all its members, are also included. The CHS covers the thirteen major metropolitan areas of Colombia, which accounted for around 45% of the country's population. It should be noted that this survey has been used for various empirical studies analyzing labor market issues in Colombia (see, for example, Magnac 1991; Attanasio, Goldberg and Pavcnik 2004; Goldberg and Pavcnik 2005). It is important to point out that, dictated by data availability, the years used are not the same for the different chapters. While chapter one uses year 2008, chapters two and three used year 2010.

This dissertation provides new relevant evidence for a developing country on three broad issues: 1) the effect of labor market segmentation in the probability that a worker is overeducated, 2) the effect of educational mismatches for explaining the difference in the returns to education for

informal and formal workers and 3) the contribution of education and informality to regional wage inequalities. This thesis is based on the following publications:

i. Herrera-Idárraga, P., López-Bazo, E. & Motellón, E. (2012) Informality and Overeducation in the Labor Market of a Developing Country. A previous version of this paper was published in the working papers series of the *Xarxa de Referència en Economia Aplicada* (XREAP working paper 2012-20). It was presented at the: Annual Conference of the European Association of Labour Economists (EALE, 2012), XXVII AIEL Conference of Labour Economics (2012), XXXVII Simposio de la Asociación Española de Economía (2012) and Second Lisbon Research Workshop on Economics, Statistics and Econometrics of Education (2013).
ii. Herrera-Idárraga, P.; López-Bazo, E.; Motellón, E. (2013) Double Penalty in Returns to Education: Informality and Educational Mismatch in the Colombian Labour market. Currently the paper is under the process of revise and resubmit in a journal listed in the ISI-JCR. A previous version of this paper was published in the working papers series of the *Research Institute of Applied Economics* (IREA Working Papers 2013/07, ISSN 2014-1254) and in the working papers series of the *Regional Quantitative Analysis Group* (AQR Working Papers 2013/04). It was presented at the: 2013 Annual Conference of the European Society for Population Economics (ESPE, 2013), X Jornadas de Economía Laboral (2013), 8th IZA/World Bank Conference on Employment and Development (2013) and 18th annual meeting of the Latin American and Caribbean Economic Association (LACEA, 2013).
iii. Herrera-Idárraga, P.; López-Bazo, E.; Motellón, E. (2014) Wage gaps across Colombian regions: the role of education and informality.

The thesis is divided into three chapters each one corresponding to the three major contributions outlined. Each chapter is organized as followed. The introduction briefly describes and motivates the topic. A section for data and descriptive analysis is presented as a preliminary evidence of the hypothesis under study. The multivariate method used to address the analysis is described in detail in another section, followed by a section that presents, analyzed and discusses the results. Finally, each chapter ends with a synthesis of the main conclusions. General conclusions, further extensions and policy recommendations of the three chapters are considered at the end of this dissertatio

Chapter 2. Informality and Overeducation in the Labor Market of a Developing Country

2.1 Introduction

There is now a substantial body of literature addressing the phenomenon of overeducation in developed countries.[1] An increasing amount of this literature is concerned with providing an explanation for overeducation that is consistent with one of the theoretical frameworks of the labor market: 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, arguing that earnings depend to some extent on both individual and job characteristics. These models also imply that there is no reason to expect wage rates to be correlated only to acquired schooling or other individual attributes (human capital theory), nor should it be expected that individual productivity and, hence, earnings will be determined solely by job characteristics (job competition model). In addition, a number of studies have also estimated the effects of overeducation on earnings. These studies show that overeducated workers tend to earn higher returns to their years of schooling than co-workers who are not overeducated, but lower returns than workers with a similar level of education who are employed in jobs that require the same level of education that they possess.

Given the differences between the labor markets of developed and developing economies, it is plausible that the factors accounting for overeducation may differ. As has already been mentioned in the introduction labors markets of developing economies are characterized by a high degree of informality. Besides the well-known negative implications of informality, primarily the result of poorer working conditions, a segmented labor market (divided between a formal and an informal jobs) might also affect the way workers match their acquired education with the education required to perform their job. As Berry and Sabot (1978) affirm, "one of the inefficiencies associated with segmentation, more difficult to document but possibly

[1]Duncan and Hoffman (1981), Verdugo and Verdugo (1989), Sicherman (1991), Tsang, Rumberger and Levin (1991), McGoldrick and Robst (1996) studied the phenomenon for the United States; Alpin, Shackleton and Walsh (1998), Green, McIntosh and Vignoles (2002), Dolton and Vignoles (2000) and Chevalier (2003) for the UK; Hartog and Oosterbeek (1998) and Groot and Massen van den Brink (2000) for Holland; Bauer (2002) and Buchel and van Ham (2003) for Germany; Kiker, Santos and De Oliveira (1997) and Mendes de Oliveira, Santos and Kiker (2000) for Portugal; Alba-Ramirez (1993) for Spain. For an extensive review of overeducation in developed countries see McGuinness (2006) and for a recent survey on overeducation see Leuven and Oosterbeek (2011).

imposing greater resource costs on the economies of developing countries, involves 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, here we assume that the study of overeducation in a developing economy with a large informal sector cannot fail to examine the role played by this segmentation.

Our assumption also builds on a model developed by Charlot and Decreuse (2005). This model shows that self-selection in education is inefficient in presence of labor market segmentation. As workers do not internalize the impact of their education decision on the others wage and employment perspectives, too many workers are willing to acquire education and this leads to overeducation. In our opinion, this is a reasonable explanation for educational mismatch in the labor markets of developing countries that presents labor market segmentation into a formal and an informal sector. In contrast with (some) developed countries in which overeducation is clearly associated with large endowments of education, the population in developing economies presents low or moderate levels of education attainment. Formal and informal labor market segmentation is, thus, a phenomenon that could account for overeducation in these economies. However, this model is not able to predict in which sector the incidence of overeducation will be highest; in this regard our empirical exercise tries to shed light on this issue.

To the best of our knowledge, few studies have examined overeducation in 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 this paucity of studies might be data limitations that hinder identification of the education levels required for specific jobs. Moreover, despite the increase in recent decades in average schooling attainment in developing countries, the average presented in these economies is lower than that presented in high-income countries. In Latin American and Caribbean Countries the average educational attainment for adult age population, 25 years and older, are 8. By comparison, the average for the OECD countries of adult age population are 10.9[2]. The fact that educational attainment remains low in developing countries means that the overeducation is a somewhat contradictory phenomenon for these economies. Nevertheless, previous studies find evidence of overeducation in

[2] This averages were computed using UNPD source:
https://data.undp.org/dataset/Mean-years-of-schooling-of-adults-years-/m67k-vi5c.

developing countries (Quinn and Rubb, 2006 for Mexico; Abbas, 2008 for Pakistan and Mehta et al., 2011 for unskilled jobs in the Philippines) and report that the incidence of overeducation is similar to that present in developed economies. For the Colombian case past studies have also found the existence of overeducation (Mora, 2005; Castillo, 2007; Dominguez Moreno, 2009[3]).

Summing up, in this chapter we study the contribution of working in a formal or an informal job on the probability of being overeducated in a developing country with low or moderate educational attainment. We hypothesize that in developing countries with a large informal sector, educated workers that do not find a high skilled formal job may accept an unskilled informal job for which she is overeducated, i.e. informal workers are more likely to be overeducated than formal workers. We test the positive relationship between informality and overeducation by exploiting information in a micro-data set for Colombian workers. In so doing, two types of empirical models are used: firstly, a simple univariate probit model that assumes that the unobservable characteristics that affect an individual's chances of working in either formal or informal jobs are independent of those determining her propensity to be overeducated; and, secondly, a bivariate probit model that enables us to control for the likely endogeneity of the selection of the formal or informal job. Our results confirm that, conditional on other individual and family characteristics, formal workers present a significantly lower probability of being overeducated. This general result seems to be driven by the fact that male informal workers face a greater probability of being overeducated, whereas no significant differences are detected between informal and formal female workers.

The remainder of this chapter is organized as follows. Next section gives the details concerning the data and presents some selected descriptors, while the empirical approach is presented in section 2.3. Section 2.4 summarizes the estimate results of the empirical models, section 2.5 presents

[3] Using micro-data for Colombia, Dominguez-Moreno (2009) studies the probability of working in the formal sector, including as an explanatory variable whether the worker is overeducated or not. In our view, this direction of the causal relationship between informality and overeducation is not correct. As a matter of fact, educational mismatch is observed after the match has happened and it is not a worker's intrinsic condition. If that were the case, studies analyzing the determinants of the probability of employment should include over-education in the list of explanatory variable. As far as we know this has not been the practice so far. On the top of that, Dominguez-Moreno (2009) did not consider the likely endogeneity of overeducation caused by the effect of omitted unobserved factors influencing both informality and overeducation.

some robustness checks dealing with the instruments and, section 2.6 presents results by gender. Finally, section 2.7 contains the conclusions.

2.2 Data and descriptive statistics

We use data from the 2008 wave of the Colombian Household Survey (CHS) for the thirteen major cities with their metropolitan areas[4]. The analysis conducted herein was limited to employed individuals between the ages of 15 and 60 that were not undertaking formal studies and who reported working more than 16 hours per week. Government employees, household employees, self-employed, bosses or employers, unpaid family workers, workers without pay in enterprises or other family businesses were not included in the sample. The subsequent sample used in the analysis comprised 15,104 observations.

As a starting point in our analysis, we had to use a criterion to determine whether a worker in the sample is overeducated, and if that worker is employed in the formal or informal sector. Four basic methods have been suggested in the literature for measuring the education required for a job and, consequently, for determining overeducation, all of which have advantages and drawbacks[5]. The first 'subjective' approach uses self-assessment to define the job's educational requirements and then compares this with the worker's actual education (Battu, Belfield and Sloane, 2000; McGuinness, 2003). The second is a variation on the above and involves asking the worker directly whether he or she is overeducated (Devillanova, 2013). The advantage of these two approaches is that they do not assign the same educational requirement to all jobs within a predetermined occupation category. However, these subjective methods might lead to biases, for examples workers may have a tendency to overestimate the requirements for performing the job, to upgrade the status of their position (Hartog, 2000).

Overeducation can also be calculated objectively by using job analysts definitions of the educational requirement for each occupation, as available in the United States Dictionary of Occupational Titles, and comparing this with the workers educational level (Rumberger, 1987; Hartog and Oosterbeek, 1988; Kiker, Santos and Mendes de Oliveira, 1997; Chevalier, 2003). Unfortunately, carrying out such analysis is very expensive, therefore this information is published only at very wide time intervals, the last updated was

[4] Bogotá, Medellin, Cali, Barranquilla, Bucaramanga, Manizales, Pasto, Pereira, Cucuta, Ibague, Monteria, Cartagena and Villavicencio. These metropolitan areas represent 45% of total population.
[5] Leuven and Oosterbeek (2011) present an extensive overview of the main drawbacks of measuring educational requirements.

carried out in 1991.

An alternative objective measure is obtained by analyzing the distribution of education in each occupation; employees who depart from the mean (Verdugo and Verdugo, 1989) or mode (Mendes de Oliveira, Santos and Kiker, 2000) are classified as being overeducated. This last approach is usually known as the 'statistical' method. It has been criticized because of the arbitrary nature of the one-standard-deviation criterion and because it might be sensitive to cohort effects. Nevertheless, it is the most common approach given that it is easy to calculate in most countries with the available data.

Since the CHS does not supply information to construct a subjective measure of overeducation, 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 those 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.[6] A worker is defined as overeducated if its education departs from the mean by more than one standard deviation. By using such an objective measure, the overall incidence of overeducation in the sample was found to be 15%, a figure similar to that reported for other developing economies (Quinn and Rubb, 2006) and lower than the incidence of overeducation in developed economies (McGuinness, 2006).

The data made available by the CHS allow us to determine whether the workers in the sample are covered or not by the social security system, and it even distinguishes 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 degree of inclusion in the social security system. Thus, we define workers as formal if they contribute both to health and old-age insurance. That is to say, an individual was classified as a formal worker if she contributed to both health and retirement pension systems. Applying this condition, as many as 33.3% of individuals in the entire

[6] As stressed in Ramos and Sanromà (2012), a two-digit classification of occupations is not optimal for applying the mode criterion. In addition, Mehta et al. (2011) emphasized that the modal education is more prone to shift even when technology and the jobs-pool do not. In any case, we also computed the results of the following sections using the mode criterion, and the main conclusions remained the same as those derived from results using the mean criteria. An appendix with these results is available on request.

sample worked in informal jobs.[7]

The incidence of overeducation for the entire sample and for formal and informal jobs is shown in the first row in Table 2.1. The percentage of formal jobs is also display in the second row. As mentioned above, 15% of Colombian urban workers were overeducated, this figure being higher in the case of formal workers (17%) than for those employed in informal jobs (11%). As for the distribution of workers in each sector, around one third had an informal job in 2008.

Differences in overeducation between the two sectors might simply be caused by disparities in the distribution of the characteristics that are assumed to affect the incidence of overeducation. Table 2.1 also displays basic summary statistics concerning the distribution of the individual and job characteristics, distinguishing between workers in the formal and informal jobs. The comparison of the figures reported in Table 2.1 confirms that there are substantial differences in some of the observable worker and job characteristics of formal and informal workers. As a matter of example, the number of 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 ability. What the figures show is that informal workers are more likely to have education levels below those of formal workers: whereas 45% of informal workers in the entire sample have at most basic secondary education, the percentage of workers in formal jobs 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 overeducation, such a gap in educational attainment could explain much of the difference observed in the overeducation figures between the two sectors.

There are significant differences in other characteristics as well. The percentage of female workers in formal jobs is higher than that in informal, 4 percentage point higher. This finding may be driving by the fact that our sample excludes self-employed individuals and unpaid family workers such as housekeepers, which concentrate a much higher proportion of female informal workers. A much larger proportion of the workforce in formal jobs is married, and workers in those jobs tend to accumulate much more tenure than informal workers, suggesting a higher stability of employment for formal workers. As for the occupational structure, the share of informal workers in

[7] Self-employment in Latin America generally constitutes one of the principle sources of employment and a large proportion of the self-employed operate in the informal sector. If the sample is not restricted to exclude self-employees, the percentage of informal workers increases up to 59% for 2008.

unskilled occupations (42%) is larger than that in the formal sector (28%). While administrative staff and professionals and technicians 1 are more strongly represented in formal jobs (24% and 9% correspondingly) than in informal jobs (14% and 2% respectively). The distribution by economic sector shows that formal workers are concentrated primarily in the industrial (25%), while there is a predominance of informal workers in sales, hotel and restaurants sector (36%). Finally, it is worth 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 of formal jobs, where more than half formal workers work in firms with more than 100 employees, and around two thirds in firms with at least 50 employees. In short, these figures indicate a close connection between informality and firm size in Colombia.

This simple descriptive analysis suggests i) the presence of quite large levels of overeducation in Colombia, ii) apparently, affecting more intensively formal workers than informal workers, and iii) that formal and informal workers differ in their levels of educational attainment, occupational distribution, and other individual and job characteristics, which are thought to exert an influence on the individual's probability of being overeducated. Since the greater incidence of overeducation in formal jobs might well be caused by a composition effect (for example, associated with the higher education of workers in that jobs), in the section that follows we estimate the impact of informality on overeducation but in relation to the conditioning factors of observable worker and job characteristics.

2.3 Empirical strategy

A multivariate empirical model needs to be specified in order to assess the impact of formal or informal sector on the probability of Colombian workers being overeducated, conditional on other observed individual, household and job characteristics. In so doing, we first assume that the allocation of a worker to a formal or informal sector is exogenous to her chances of being overeducated. Under such an assumption, a univariate probabilistic specification provides consistent estimates of the effect of the sector on the chances of the worker having more education than that required for her occupation. However, the endogeneity assumption can easily be questioned. Were this to be the case, the standard probabilistic specification with exogenous covariates would lack consistency. To address this issue, we

estimate the effect of the sector by means of a bivariate specification in which this variable is instrumented.

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

$$Y_i^* = \beta X_i + \alpha S_i + \varepsilon_i \tag{2.1}$$

where X_i is a vector of individual and firm characteristics (including age, gender, marital status, head of household, education, tenure, occupation, industry sector 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 overeducated $(Y_i^* \geq 0)$
$Y_i = 0$ otherwise

Given the normality of the error term in eq. (2.1) a probit specification can be used to estimate the effect of the sector on the probability of being overeducated, conditional on the other characteristics in *X*:

$$P[Y_i = 1] = P[\beta X_i + \alpha S_i + \varepsilon_i > 0] = \Phi[\beta X_i + \alpha S_i] \tag{2.2}$$

where Φ denotes the standard normal cumulative distribution function.

Since the estimate of the coefficient α is only informative about the sign of the impact of *S*, its associated marginal effect is computed from the estimates of the probit model in eq. (2.2) as:

$$\partial P[(Y = 1)/S]_{\bar{X}} = \Phi(\beta \bar{X} + \alpha) - \Phi(\beta \bar{X}) \tag{2.3}$$

where the bar over the X denotes the sample average.

As indicated above, the assumption made in the specification of the univariate probit in eq. (2.2) is that the sector (formal or informal) is exogenous to the probability of being overeducated. However, if the assignment of workers to each of the sectors is not random and some

unobservable factors (ability among others) that influence the probability of being assigned to a particular sector are also affecting the probability of being overeducated, then the estimation of a univariate probit would suffer from selection bias.[8] This would have dramatic consequences on the inference since the estimates from the univariate probit would be inconsistent if this endogeneity was ignored.

To take account of this potential drawback properly, in a second step, we estimate the effect of the sector in a bivariate probit model, in which the sector is instrumented by family characteristics. In addition to the latent outcome equation in (2.1), the bivariate model is based on an additional equation for the latent model linking the probability of assignment to the formal or informal sector to a set of characteristics:

$$S_i^* = \gamma Z_i + \mu_i \tag{2.4}$$

where Z_i is a vector of observed individual and family characteristics, and μ_i is a normally distributed error term. Z_i includes the set of characteristics in X_i plus some additional variables used as instruments for the assignment to the sector, S_i^*. Since we can only observe the sector 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 is formal $(S_i^* \geq 0)$
$S_i = 0$ otherwise

Therefore, the probit specification associated with the probability of working in a formal job, conditioned to the characteristics in Z, stands as:

$$P[S_i = 1] = P[\gamma Z_i + \mu_i > 0] = \Phi[\gamma Z_i] \tag{2.5}$$

The bivariate probit thus consists of equations (2.2) and (2.5), 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 words, the empirical model allows for the likely

[8] We have ignored another type of selection whereby an individual might not accept a job that does not match his or her level of education and chooses instead to be unemployed or to remain outside the labor force. We argue that this selection bias is irrelevant in the case of Colombia where there is no unemployment benefit system and the family protection network against unemployment is low or exclusive to a group of high-income individuals.

correlation of the unobserved determinants of overeducation and the unobserved determinants of the sector. 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 (L) associated to this set of events is (for further details see Wooldridge 2002, p.478):

$$L = \sum_{Y_i=1,S_i=1} ln\Phi_2[\beta X_i + \alpha S_i, \gamma Z_i, \rho] + \sum_{Y_i=1,S_i=0} ln\Phi_2[\beta X_i, -\gamma Z_i, -\rho] + \sum_{Y_i=0,S_i=1} ln\Phi_2[-\beta X_i - \alpha S_i, \gamma Z_i, -\rho] + \sum_{Y_i=0,S_i=0} ln\Phi_2[-\beta X_i, -\gamma Z_i, \rho] \quad (2.6)$$

The inference in the bivariate probit model is based on the maximization of the log-likelihood in eq. (2.6) with respect to the parameters β, α, γ and ρ. If ρ is statistically different from 0, the endogeneity of the assignment to the formal or the informal sector would be confirmed, and thus estimates from the bivariate probit are preferable; otherwise conclusions regarding the impact of the sector could be based on the estimate of the univariate probit in eq. (2.2).[9] As in the case of the univariate probit model, marginal effects are computed from the estimates of the bivariate probit model to assess the contribution of each variable to the probability of being overeducated.

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 relying solely on the functional form and the distributional assumptions. However, the objective of forming a consistent estimator for α becomes manageable if we can construct at least one 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 overeducation equation (outcome equation). The first condition is easy to check; 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. In the context of the bivariate probit model,

[9] A bivariate probit model with an endogenous binary regressor has been used in, for instance, Evans and Schwab (1995) to analyze the effect of catholic schools on finishing high school and starting college.

this condition relies on the economic or institutional knowledge related to the problem under study.

As in many other studies, finding suitable instrumental variables is far from straightforward, since almost any regressor that determines the probability of being overeducated could plausibly affect assignment into formal and informal jobs as well. Previews studies about informality control for household characteristics, that may affect a person's propensity to be employed in the informal sector, such as the number of children in a household, number of inactive adults in a household, and earnings of other household members (Hill, 1983; Magnac, 1991, Marcouiller, Ruiz de Castilla and Woodruff, 1997; Goldberg and Pavcnik, 2003; Pisani and Pagán, 2004; Maloney, 2004). To the best of our knowledge, in the over-education literature only Mavromaras and McGuinness (2012) use the presence of children as a control variable in probit estimations of overskilling, situation where a worker reports that their skills are not fully utilized in their job. The authors only report a marginal statistical significance for the coefficient of this variable, and only for the group of moderately overskilled workers. Thus, it could be the case that certain family characteristics influence an individual's choice regarding formal or informal employment but do not affect overeducation, such as the presence of children in the household and the earnings of other household members. One reason why such family characteristics may affect the sector of employment is because they are closely related to the households income needs. For instance, having more children means more expenses for the household and increase the need of finding a job, which is presumably more easily available in the informal sector. The assumption here is that the presence of children does not exert a significant effect on the propensity to be overeducated. Another family characteristic that is thought to influence the choice of employment sector but not the individual's propensity to be overeducated is the social status, which we suggest is captured by the educational achievement of other members of the household. Accordingly, we construct the average number of years of schooling of other household members and we used it as an additional instrument for the sector of employment. Table A2.1 presents descriptive statistics of the instruments. As it can be seen, informal workers are more likely to live in households with lower levels of education of its members and tend to live in households with presence of children (aged between 0 and 8).

2.4 Results

2.4.1 Probit results

The maximum likelihood estimates of the coefficients when running the univariate probit model eq. (2.2) are reported in Table 2.2. The corresponding marginal effects for the average individual as defined in eq. (2.3) are also reported. Our results show that after controlling for other characteristics, formal workers are less likely to be overeducated than their informal counterparts. In other words, when we compare formal and informal workers with similar individual, household, and firm characteristics, those in the former group have a lower propensity to be overeducated. This contrasts sharply with the raw probabilities derived from the sample since, as the descriptive analysis shows, the share of overeducated workers in the formal sector is greater than that in the informal sector. Thus, these results suggest that a sorting effect drives the gap in the raw propensities.

Yet, it should be mentioned that the marginal effect associated with working in formal jobs is of a moderate magnitude. The probability that a formal worker is overeducated is just 2.5 percentage points (pp) less than that for an otherwise similar informal worker. Thus, the results from the univariate probit model suggest a modest impact of formality on overeducation having first controlled for education and other observable characteristics.

In the case of the estimates of the coefficients for the control variables, the results shown in Table 2.2 are consistent with previous findings in the literature. As expected, the probability of being overeducated increases with educational attainment (Alba-Ramirez, 1993; Kiker, Santos and Mendes de Oliveira, 1997; Quinn and Rubb, 2006). Overeducated workers may substitute education for a lack of job experience, taking jobs that require less education than they actually possess in order to accumulate experience and improve their chances of finding a better job match (Rosen, 1972; Sicherman and Galor, 1990; Mendes de Oliveira, Santos and Kiker, 2000). To test this hypothesis we use a variable that measures experience, specifically potential experience calculated as an individual's age minus years of education minus 5 years (in Colombia, children start attending to school at age of 5). On the other hand, several studies report that overeducation may have a negative effect on job satisfaction (Tsang, Rumberger and Levin, 1991), if this is the case, then overeducated workers with more tenure in a firm can be expected to be more prone to turnover. Consequently we hypothesize that overeducated workers

will have less tenure. The results for the estimated marginal effect of general experience (years since leaving the education system) confirm the expected negative effect of this variable on the probability of an average worker in the sample being overeducated. However, it should be pointed out that this marginal effect is only significantly different from zero at a 10% confidence level. The impact of tenure is also negative, though almost negligible and, not in fact statistically significant. Therefore, results for Colombia are not conclusive regarding the evidence on the substitutability between education and other forms of human capital postulated by the human capital theory, according to which overeducation might be seen as a transitory situation.

The results also indicate that females are less likely to be overeducated than males presenting similar characteristics, and that marital status does not have a statistically significant impact on the probability of being overeducated for both genders. Significant differences do exist however in terms of industry and firm size. Compared to individuals employed in Agriculture, mining, electricity, gas and water (our reference category), those employed in construction are more likely to be overeducated (8.4pp), while those working in transportation, financial intermediation and social services are less likely to be overeducated. As for firm size, compared to working in a micro (1 to 10 workers), working in a small firm (11 to 50 workers) or in a large firm (101 workers or more) does not seem to have a significant impact on the probability of being overeducated. While medium size firms (51 to 100 workers) have a negative and significant effect on the likelihood of being overeducated. It is worth mentioning that local labor market conditions do not seem to be relevant, as the coefficient of the metropolitan unemployment rate, although positive, is not statistically significant. Finally, individual characteristics such as being the head of the household or being married do not have a significant impact on the probability of being overeducated. Being a woman reduces the probability of being overeducated by 4pp, however being a women head of household increase the probability by 3pp.

Nonetheless, it should be borne in mind that the specification used to obtain these results assumes the exogeneity of the employment and the absence of a simultaneous impact of the unobservable characteristics on the probability of overeducation and on the assignment of formal or informal sector. The violation of these assumptions would invalidate the results.

2.4.2 Biprobit results

Our estimates of the effect of the sector, when relaxing the assumption of exogeneity and the lack of correlation between the unobservable variables that influence both overeducation and formality/informality, are summarized in Table 2.3. These results correspond to the maximum likelihood estimates obtained from the bivariate probit model eq. 2.6 described in section 3, using instruments for the employment sector and the same set of control variables as those employed in the univariate probit model. Here, the discussion focuses solely on the coefficients of the equation for the probability of being overeducated since the estimates obtained for the parameters in the formal/informal sector equation (see Table A2.2 in the appendix) are relatively standard, and largely conform to results reported elsewhere (Magnac, 1991; Pradhan and van Soest, 1995).

The coefficient of the formal sector and the corresponding marginal effect are estimated to be negative and highly significant. In fact, the magnitude of the marginal effect of working in the formal sector estimated from the bivariate probit model is substantially higher than that estimated by the univariate probit model. The results suggest that, for otherwise similar workers, working in the formal sector reduces the probability of overeducation by 16.44 pp. This finding confirms that selection bias strongly affects the estimate of the effect of the employment sector on the probability of being overeducated and, hence, the need to account for it. On the other hand, it seems that, in addition to the benefits associated with receiving social security and higher wages, being a formal worker also ensures a better use of one's skills in the workplace. Or, alternatively, an informal worker besides not having social security and receiving lower wages, as suggested by previews literature, is less likely to make proper use of his acquired education in his job. As discussed in the introduction, to the best of our knowledge this finding has not previously been recorded, and represents a novel contribution of this study.

Note that the estimate of ρ (correlation between the error terms of the overeducation and the employment sector equations) is positive and statistically significant, suggesting that non-observable characteristics that exert a positive effect on the probability of being formal employed also have a positive impact on the probability of being overeducated. This could be interpreted as evidence that in the case of formal workers overeducation is caused, to some extent, by the desire to form part of the formal sector (better

employment opportunities, social system protection, etc.). A worker with a certain level of education might take a job for which less education is actually required, simply because that job is protected, for example, by the minimum wage.

An alternative interpretation of the positive effect of unobservable factors on the probability of being overeducated can be made from within an internal labor market framework (Doeringer and Piore, 1972). Internal labor markets are those in which workers are hired into entry-level jobs, while higher levels are filled from within. Certain rules differentiate the 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 port-of-entry jobs, presumably at low levels, that are filled through external hiring. In this context, individuals in any given firm are hired into its lower or middle levels and subsequently succeed in advancing to higher levels. Workers that do not have the qualifications for particular entry-level jobs are thus excluded from accessing the entire job ladder. For this reason, workers may initially accept a job for which their actual education is higher than that actually required in exchange for the benefits of gaining access to an internal labor market. It should be stressed that internal labor markets operate in the primary sector (formal) rather than in the secondary sector (informal).

As for the estimate of the coefficients, and the associated marginal effects of the other observable characteristics in the overeducation equation, they are, in general, roughly identical to those estimated with the univariate probit, with the exception of firm size. The estimates from the bivariate probit model indicate that compared to individuals working for micro-firms (those with less than 10 workers), workers in small, medium and large firms are more likely to be overeducated. This result can be interpreted as follows: large firms usually have better job opportunities (as well as paying higher wages), and workers have better chances of being promoted and of receiving more on-the-job training. These characteristics mean that job offers from large firms are valued highly by job seekers who might apply for vacancies in which the required level of education is less than the one they have acquired. Likewise, large firms in the formal sector are in a position to select the most highly skilled from the pool of available workers. Yet, it should be pointed out that

the impact on overeducation is weaker in the case of medium size firms (between 50 and 100 workers), where the coefficient is not, in fact, statistically significant.

2.5 Validity of the instruments

The estimates of the bivariate probit presented in Table 2.3 are consistent and unbiased as long as the instruments are correlated with the probability of working in formal or informal sector but not with the error term of the overeducation equation in (2.2). In order to investigate if the selected instruments are valid we implement a procedure suggested by Cohen-Zadar and Elder (2009) and also implemented by Kim (2011). This approach is based on the idea that the instruments, presence of children and average years of education of the other members of the household, exert an effect on the probability of overeducation only through the sector, if it is formal or informal, but not directly. If the instruments do not influence the probability of overeducation apart from its effect on the sector, it should have no effect in the overeducation in a subsample of workers for whom the probability of working in either informal or formal jobs is closely to zero.

One can argue that public employees are a specific group of workers for whom the probability of working in informal jobs is approximately zero.[10] Then, for this subsample of workers, the instruments should have no effect in the probability of being overeducated. Table 2.4 reports the effects of the educational achievement of other members of the household and the presence of children estimated from a probit overeducation equation for public employees, conditioning on the other set of controls used for the estimates of the probit overeducation equation for private employees in Table 2.3. Results in Table 2.4, for workers working in the public sector, confirm that the coefficients of the instruments are not statistically significant, which means that the instruments do not exert a direct effect on the probability of being overeducated. Although an insignificant estimate for the coefficients associated to these variables are not guarantee of exogeneity, it does provide some evidence that their use as instruments is likely not to be problematic.

Last as a sensitive analysis for the biprobit estimates, we estimate the effect of the sector of employment on the probability of being overeducated using different set of instruments. This sensitive analysis is presented in Table

[10] As a matter of fact, only 3.9% of the public employees report that they don't make contribution to the health and old insurance system in contrast with the 33% of workers from private firms.

2.5. For simplicity, and for purposes of comparison, the estimates of Table 2.3 are presented in column [1]. The results of the biprobit when using only the average years of education of the other members of the household as an instrument are summarized in column [2], whereas those using only the presence of children as an instrument are shown in column [3]. As it can be seen the results reported in Table 2.5 show that the estimated effect of the sector of employment on the probability of being overeducated is fairly robust to the set of instrument chosen. Still, the effect of the sector of employment is estimated to be lower when using as instruments only the dummies for the presence of children.

2.6 Results by gender

To obtain some insights into differences by gender and how sensible our results are to this dimension all the results were computed for men and women separately. As mentioned above, 15% of Colombian urban workers were overeducated, this figure being higher in the case of formal workers (17%) than for those employed in informal jobs (12%). Table 2.5 shows that this gap of six percentage points is also found for both male and female workers. As for the distribution of workers in each sector, around one third had an informal job in 2008, this percentage being higher for men (35%) than for women (31%).

Table 2.5 also shows that male and female workers differ in some of the characteristics that are supposed to affect overeducation. Interestingly, the most remarkable differences are those of the distribution of education levels and occupations. Broadly speaking, female workers are more highly educated than their male counterparts, and find themselves concentrated in occupations such as administrative staff (24%), merchant and vendor jobs (22%) and service work (20%), while men are more highly concentrated in unskilled occupations (48%), which are associated with higher levels of informality. As for industry sector, women are more concentrated in social services (26%) while men tend to be more concentrated in industry (26%)

These differences in characteristics for men and women may affect the results that have been presented so far. Since we found that the results for the total sample were sensitive to the problem of endogeneity of the sector, the discussion of the results for men and women will focus on the estimates of the biprobit (Table A2.3 in the appendix presents the results for the univariate probit). Results by gender for the biprobit model presented in Table 2.7 point

to a substantial gender differences in the impact of the sector on the probability of being overeducated. Whereas, for a male, having a formal job reduces the propensity of overeducation by 20.09 pp compared to a similar informal male worker, for females the effect is lower, 10.71 pp and it is statistical significant only at the 5% confidence level. Interestingly, for the female workers we do not find a significant correlation between the errors of the two equations in contrast with the highly significant correlation coefficient for males.

As for the validity of the instruments, Table 2.8 confirms that for female public employees the coefficients of the instruments are not statistically significant, which means that the instruments do not exert a direct effect on the probability of being overeducated. In the case of male public employees the educational achievement of other members of the household is statistical significant only at 5% but its marginal effect is considerably low -0.0075.

Finally Table 2.9 presents the sensitive analysis for the biprobit estimates when using different set of instruments for men and women. It was found that the effect of the sector of employment is estimated to be lower when using as instruments only the dummies for the presence of children in the case of men. While for the sample of female workers, none of the two types of instruments provides with a significant estimate of the effect of the sector of employment.

Results by gender show that there are important differences in the probability of being overeducated between men and women. While informal male workers are more likely to be overeducated, the propensity of being overeducated among women does not seem to be closely related to the sector choice.

2.7 Conclusions

This study has sought to add to the overeducation literature by analyzing the connection between labor market segmentation, into a formal and informal sector, and overeducation in a developing country. To date, studies concerned with informality in developing countries have focused primarily on the size of this sector, on the effects of labor market rigidities on employment, wages and their distribution, and on the probability of a worker entering the informal sector. However, no attention has been paid to the effects that a large informal sector has on the way workers match their

education with that required performing their particular jobs. This study offers some new evidence in this respect.

Using micro-data for Colombia, we have estimated two types of empirical models in order to test the relationship between overeducation and informality. A simple univariate probit model for the probability of being overeducated that includes the sector in which the individual is employed as an explanatory factor, formal or informal. And a bivariate probit model with an endogenous regressor that considers that the assignment of workers to each of the sector is not random and some unobservable factors that influence the probability of choosing a particular sector, could also affect the probability of being overeducated. The results of the univariate probit estimation indicate that formal workers are less likely to be overeducated than their informal counterparts. However, we have also shown that the assignment of workers to the formal or informal jobs is not random and that some unobservable characteristics that influence the probability of choosing a particular sector also affect the probability of being overeducated.

The results obtained from the bivariate probit model for the probability of overeducation, once the potential endogeneity of sector choice and overeducation were taken into account, confirm that formal workers are less likely to be overeducated and that non-observable characteristics that exert a positive effect on the probability of being a formal worker have a positive impact on the probability of being overeducated. This could be interpreted as evidence that for formal male workers; overeducation is caused, at least in part, by a desire to have a formal job (better employment opportunities, social system protection, etc.). A worker with a good education may take a job for which less education is required, because that job is protected, for example, by the minimum wage. This negative effect of formality over the probability of being overeducated was found to be relevant for the case of male workers but not as much for females.

According to our results it seems that, in addition to the benefits associated with receiving social security and earning higher wages, being a formal worker also ensures a better use of acquired skills in the workplace. To the best of our knowledge, no study has presented evidence of this to date.

Table 2.1. Descriptive statistics for the main variables in the analysis

Variable	Total	Formal	Informal
Overeducation	0.15	0.17	0.12
Informal	0.33	-	-
Age (years)	33.93	34.69	32.38
Experience (years)	17.97	17.85	18.23
Tenure (months)	48.56	59.00	27.51
Women	0.43	0.44	0.40
Married	0.53	0.55	0.48
Household Head	0.40	0.41	0.37
Educational Attainment			
Basic Primary or below	0.13	0.08	0.23
Basic secondary	0.14	0.10	0.22
Secondary	0.36	0.36	0.37
Higher education or more	0.36	0.45	0.18
Education (years)	10.96	11.85	9.16
Occupation			
Unskilled	0.33	0.28	0.42
Professionals and Technicians 1	0.07	0.09	0.02
Professionals and Technicians 2	0.05	0.05	0.04
Managers and Public Officials	0.03	0.04	0.02
Administrative Staff	0.20	0.24	0.14
Merchant and Vendor	0.17	0.16	0.18
Service Worker	0.16	0.15	0.18
Firm size			
Micro (1 -10 workers)	0.32	0.15	0.68
Small (11 - 50 workers)	0.22	0.24	0.18
Medium (51 - 100 workers)	0.07	0.09	0.03
Large (101 workers or more)	0.39	0.53	0.11
Sector			
Mining, electricity, gas and water	0.03	0.04	0.01
Industry	0.24	0.25	0.21
Construction	0.08	0.04	0.14
Sales, Hotels and Restaurants	0.28	0.24	0.36
Transportation	0.08	0.09	0.07
Financial Intermediation	0.12	0.14	0.07
Social Services	0.17	0.20	0.12
Observations	15104	10098	5006

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

Table 2.2. Estimates from the univariate probit over-education model

	Coefficient	Marginal Effect
Formal	-0.1498**	-0.0250**
	[0.0384]	[0.0065]
Schooling years	0.2409**	0.0401**
	[0.0056]	[0.0012]
Experience (years)	0.0096+	-0.0006+
	[0.0056]	[0.0003]
Experience^2	-0.0004*	
	[0.0001]	
Tenure (months)	-0.0006	-0.0001
	[0.0006]	[0.0001]
Tenure^2	0.0000	
	[0.0000]	
Women	-0.2398**	-0.0400**
	[0.0419]	[0.0070]
Married	-0.0232	-0.0039
	[0.0493]	[0.0082]
Women Married	-0.001	-0.0002
	[0.0661]	[0.0110]
Household head	-0.049	-0.0082
	[0.0493]	[0.0082]
Women Household head	0.1782*	0.0297*
	[0.0712]	[0.0119]
Industry	0.0737	0.0123
	[0.0845]	[0.0141]
Construction	0.5034**	0.0839**
	[0.0983]	[0.0164]
Sales, Hotels, Restaurants	-0.1029	-0.0171
	[0.0848]	[0.0142]
Transportation	-0.3531**	-0.0588**
	[0.0928]	[0.0156]
Financial Intermediation	-0.4160**	-0.0693**
	[0.0895]	[0.0150]
Social Services	-0.4907**	-0.0818**
	[0.0879]	[0.0147]
Firm Size Small	-0.0271	-0.0045
	[0.0428]	[0.0071]
Firm Size Medium	-0.2150**	-0.0358**
	[0.0662]	[0.0110]
Firm Size Large	0.0022	0.0004
	[0.0414]	[0.0069]
Metro. Area U. Rate	0.0054	0.0009
	[0.0078]	[0.0013]
Constant	-3.6861**	
	[0.1484]	
Observations	15675	
Log pseudolikelihood	-5242.92	

Notes: Robust standard errors in [].+ p<0.1, * p<0.05, ** p<0.01. Marginal effects for experience and tenure are calculated using the coefficient of the linear and quadratic term.

Table 2.3. Estimates from the bivariate probit model for the overeducation equation

	Coefficient	Marginal Effect
Formal	-0.9075**	-0.1644**
	[0.1760]	[0.0380]
Schooling years	0.2509**	0.0454**
	[0.0056]	[0.0021]
Experience (years)	0.0122*	-0.0002*
	[0.0056]	[0.0003]
Experience^2	-0.0004**	
	[0.0001]	
Tenure (months)	0.0007	0.0001
	[0.0006]	[0.0001]
Tenure^2	0.000	
	[0.0000]	
Women	-0.2013**	-0.0365**
	[0.0421]	[0.0075]
Married	0.0278	0.005
	[0.0525]	[0.0095]
Women Married	-0.0614	-0.0111
	[0.0682]	[0.0124]
Household head	-0.0661	-0.012
	[0.0520]	[0.0094]
Women Household head	0.1967*	0.0356*
	[0.0765]	[0.0137]
Industry	0.0362	0.0066
	[0.0847]	[0.0153]
Construction	0.3834**	0.0694**
	[0.1026]	[0.0179]
Sales, Hotels, Restaurants	-0.1266	-0.0229
	[0.0846]	[0.0154]
Transportation	-0.3724**	-0.0674**
	[0.0924]	[0.0171]
Financial Intermediation	-0.4038**	-0.0731**
	[0.0903]	[0.0163]
Social Services	-0.5382**	-0.0975**
	[0.0881]	[0.0166]
Firm Size Small	0.2540**	0.0460**
	[0.0765]	[0.0154]
Firm Size Medium	0.1438	0.0261
	[0.1029]	[0.0194]
Firm Size Large	0.3624**	0.0656**
	[0.0901]	[0.0186]
Metro. Area U. Rate	-0.0005	-0.0001
	[0.0079]	[0.0014]
Constant	-3.4673**	
	[0.1702]	
ρ	0.4347**	
	[0.1211]	
Observations	15104	
Log pseudolikelihood	-11384.32	

Notes: Robust standard errors in [].+ p<0.1, * p<0.05, ** p<0.01. Marginal effects for experience and tenure are calculated using the coefficient of the linear and quadratic term.

Table 2.4. Reduced-form relationship between family characteristics and overeducation probability among public employees

	Coefficient	Marginal Effect
Average years of education other members	-0.0169	-0.0036
	[0.0117]	[0.0025]
Number of kids age 0	0.2367	0.051
	[0.1832]	[0.0395]
Number of kids age 1	0.0261	0.0056
	[0.2318]	[0.0500]
Number of kids age 2	0.1287	0.0278
	[0.2065]	[0.0445]
Number of kids age 3	-0.0103	-0.0022
	[0.1643]	[0.0354]
Number of kids age 4	0.105	0.0226
	[0.1679]	[0.0362]
Number of kids age 5	0.1433	0.0309
	[0.1860]	[0.0401]
Number of kids age 6	-0.0861	-0.0186
	[0.1948]	[0.0420]
Number of kids age 7	-0.0882	-0.019
	[0.1528]	[0.0330]
Number of kids age 8	0.0324	0.007
	[0.1690]	[0.0365]
Wald Test - Joint Significance	χ^2	p-value
Average years of educ.	2.07	0.1501
Num. of kids age 0 - 8	3.48	0.9421
All instruments	5.73	0.8377
Observations	1823	
Log pseudolikelihood	-691.89739	

Notes: Robust standard errors in [].+ p<0.1, * p<0.05, ** p<0.01. Other explanatory variables, except the sector of employment, listed in Table 2.2 are also included in the regression

Table 2.5. Estimates of the sector of employment on overeducation with different set of instruments

	Biprobit - IV		
	[1]	[2]	[3]
Formal	-0.1644**	-0.1581**	-0.1151**
	[0.0380]	[0.0438]	[0.0434]
ρ	0.4657**	0.4406**	0.2991*
	[0.1211]	[0.1395]	[0.1390]
Instruments			
Average years of education other members	Yes	Yes	
Num. Children 0 - 8 years old	Yes		Yes

Notes: Robust standard errors are in []. + p<0.1, * p<0.05, ** p<0.01. Marginal effects are presented. The exogenous variables of individual's characteristics, job's characteristics and the unemployment rate of the metropolitan listed in Table 2.2 are included in all regressions.

Table 2.6. Descriptive statistics for the main variables in the analysis for men and women

	Men			Women		
Variable	Total	Formal	Informal	Total	Formal	Informal
Overeducation	0.16	0.18	0.12	0.15	0.17	0.11
Informal	0.35	-	-	0.31	-	-
Age (years)	34.09	35.11	32.19	33.71	34.17	32.67
Experience (years)	18.77	18.84	18.64	16.91	16.60	17.60
Tenure (months)	48.69	59.84	27.92	48.39	57.94	26.90
Married	0.61	0.65	0.54	0.41	0.42	0.39
Household Head	0.54	0.58	0.47	0.21	0.21	0.23
Educational Attainment						
Basic Primary or below	0.17	0.11	0.28	0.09	0.05	0.16
Basic secondary	0.17	0.13	0.25	0.10	0.07	0.17
Secondary	0.38	0.39	0.34	0.35	0.32	0.41
Higher education or more	0.28	0.37	0.13	0.46	0.56	0.26
Education (years)	10.32	11.27	8.55	11.80	12.57	10.07
Occupation						
Unskilled	0.48	0.41	0.60	0.13	0.12	0.15
Professionals and Technicians 1	0.07	0.09	0.02	0.06	0.09	0.02
Professionals and Technicians 2	0.03	0.04	0.03	0.07	0.07	0.06
Managers and Public Officials	0.03	0.04	0.01	0.04	0.04	0.03
Administrative Staff	0.14	0.16	0.10	0.28	0.33	0.19
Merchant and Vendor	0.12	0.12	0.13	0.22	0.21	0.26
Service Worker	0.13	0.14	0.10	0.20	0.15	0.29
Firm size						
Micro (1 -10 workers)	0.33	0.14	0.68	0.31	0.16	0.67
Small (11 - 50 workers)	0.22	0.24	0.18	0.22	0.23	0.18
Medium (51 - 100 workers)	0.07	0.09	0.04	0.06	0.08	0.03
Large (101 workers or more)	0.38	0.53	0.10	0.40	0.53	0.13
Sector						
Mining, electricity, gas and water	0.04	0.05	0.02	0.02	0.02	0.01
Industry	0.26	0.28	0.23	0.20	0.20	0.19
Construction	0.12	0.07	0.23	0.01	0.01	0.01
Sales, Hotels and Restaurants	0.26	0.24	0.31	0.31	0.26	0.44
Transportation	0.10	0.10	0.08	0.07	0.07	0.07
Financial Intermediation	0.12	0.14	0.08	0.13	0.15	0.07
Social Services	0.10	0.12	0.06	0.26	0.29	0.21
Observations	8629	5616	3013	6475	4482	1993

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

Table 2.7. Estimates from the bivariate probit model for the overeducation equation for men and women

	Men		Women	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Formal	-1.1884**	-0.2009**	-0.5617*	-0.1071*
	[0.2030]	[0.0468]	[0.2242]	[0.0459]
Schooling years	0.2750**	0.0465**	0.2250**	0.0429**
	[0.0080]	[0.0030]	[0.0094]	[0.0026]
Experience (years)	0.0079	-0.0009	0.0171*	0.0011*
	[0.0078]	[0.0004]	[0.0083]	[0.0006]
Experience^2	-0.0004+		-0.0003	
	[0.0002]		[0.0002]	
Tenure (months)	0.0009	0.0001	0.0004	0.0000
	[0.0008]	[0.0001]	[0.0011]	[0.0001]
Tenure^2	0.0000		0.0000	
	[0.0000]		[0.0000]	
Married	0.0647	0.0109	-0.0566	-0.0108
	[0.0541]	[0.0092]	[0.0462]	[0.0089]
Household head	-0.0466	-0.0079	0.0792	0.0151
	[0.0537]	[0.0091]	[0.0586]	[0.0111]
Industry	0.0902	0.0153	-0.0577	-0.011
	[0.1046]	[0.0176]	[0.1511]	[0.0289]
Construction	0.4043**	0.0683**	-0.1085	-0.0207
	[0.1249]	[0.0199]	[0.2180]	[0.0416]
Sales, Hotels, Restaurants	-0.1639	-0.0277	-0.0343	-0.0065
	[0.1057]	[0.0180]	[0.1491]	[0.0285]
Transportation	-0.4605**	-0.0778**	-0.2033	-0.0388
	[0.1177]	[0.0205]	[0.1593]	[0.0305]
Financial Intermediation	-0.4760**	-0.0805**	-0.2901+	-0.0553+
	[0.1164]	[0.0196]	[0.1539]	[0.0293]
Social Services	-0.8044**	-0.1360**	-0.3604*	-0.0687*
	[0.1196]	[0.0217]	[0.1491]	[0.0286]
Firm Size Small	0.3062**	0.0518**	0.2190*	0.0418*
	[0.0982]	[0.0194]	[0.0986]	[0.0198]
Firm Size Medium	0.2765*	0.0467*	0.0326	0.0062
	[0.1288]	[0.0241]	[0.1383]	[0.0265]
Firm Size Large	0.4151**	0.0702**	0.3382**	0.0645**
	[0.1147]	[0.0233]	[0.1129]	[0.0232]
Metro. Area U. Rate	-0.0051	-0.0009	0.0062	0.0012
	[0.0106]	[0.0018]	[0.0118]	[0.0022]
Constant	-3.4839**		-3.7693**	
	[0.2268]		[0.2544]	
ρ	0.5987**		0.2424+	
	[0.1670]		[0.1331]	
Observations	8629		6475	
Log pseudolikelihood	-6346.62		-4952.43	

Notes: Robust standard errors in [].+ p<0.1, * p<0.05, ** p<0.01. Marginal effects for experience and tenure are calculated using the coefficient of the linear and quadratic term.

Table 2.8. Reduced-form relationship between family characteristics and overeducation probability among public employees for men and women

	Men		Women	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Average years of education other members	-0.0385*	-0.0075*	-0.0057	-0.0013
	[0.0192]	[0.0037]	[0.0155]	[0.0035]
Number of kids age 0	0.226	0.0438	0.3455	0.0773
	[0.2525]	[0.0490]	[0.2823]	[0.0631]
Number of kids age 1	-0.1384	-0.0268	0.3952	0.0884
	[0.3191]	[0.0617]	[0.3319]	[0.0741]
Number of kids age 2	0.1939	0.0376	0.0395	0.0088
	[0.3155]	[0.0614]	[0.2791]	[0.0625]
Number of kids age 3	0.1534	0.0297	-0.2101	-0.047
	[0.2066]	[0.0398]	[0.2831]	[0.0634]
Number of kids age 4	0.3126	0.0605	-0.49	-0.1097
	[0.2156]	[0.0415]	[0.3222]	[0.0721]
Number of kids age 5	0.0563	0.0109	0.2933	0.0656
	[0.2602]	[0.0504]	[0.2904]	[0.0649]
Number of kids age 6	-0.0475	-0.0092	-0.2374	-0.0531
	[0.2894]	[0.0560]	[0.2814]	[0.0632]
Number of kids age 7	-0.1797	-0.0348	0.0336	0.0075
	[0.2163]	[0.0418]	[0.2103]	[0.0471]
Number of kids age 8	-0.3145	-0.0609	0.2897	0.0648
	[0.2365]	[0.0457]	[0.2189]	[0.0489]
Wald Test - Joint Significance	χ^2	p-value	χ^2	p-value
Average years of educ.	4.01	0.0451	0.14	0.7115
Num. of kids age 0 - 8	6.85	0.6531	9.54	0.3892
All instruments	11.52	0.3182	9.76	0.4617
Observations	882		938	
Log pseudolikelihood	-304.19913		-368.02025	

Notes: Robust standard errors in [].+ p<0.1, * p<0.05, ** p<0.01. Other explanatory variables, except the sector of employment, listed in Table 2.2 are also included in the regression.

Table 2.9. Estimates of the sector of employment on overeducation with different set of instruments for men and women

Men	Biprobit - IV		
	[1]	[2]	[3]
Formal job	-0.2009**	-0.2175**	-0.1658**
	[0.0468]	[0.0459]	[0.0718]
ρ	0.6911**	0.7440**	0.5529*
	[0.1670]	[0.1642]	[0.2527]
Women	Biprobit - IV		
	[1]	[2]	[3]
Formal job	-0.1071*	-0.0616	-0.0749+
	[0.0459]	[0.0444]	[0.0419]
ρ	0.2473+	0.109	0.1455
	[0.1331]	[0.1322]	[0.1220]
Instruments			
Average years of education other members	Yes	Yes	
Num. Children 0 - 8 years old	Yes		Yes

Notes: Robust standard errors are in []. + p<0.1, * p<0.05, ** p<0.01. Marginal effects are presented. The exogenous variables of individual's characteristics, job's characteristics and the unemployment rate of the metropolitan listed in Table 2.2 are included in all regressions.

.

Appendix

Table A2.1. Descriptive statistics of household characteristics

	Total	Informal	Formal
Average years of education other members	8.85	7.70	9.42
Number of kids age 0	0.07	0.09	0.06
Number of kids age 1	0.08	0.09	0.07
Number of kids age 2	0.08	0.09	0.07
Number of kids age 3	0.07	0.09	0.07
Number of kids age 4	0.07	0.09	0.07
Number of kids age 5	0.08	0.09	0.07
Number of kids age 6	0.07	0.09	0.07
Number of kids age 7	0.08	0.09	0.07
Number of kids age 8	0.08	0.10	0.07
Observations	15104	5006	10098

Notes: Figures are percentage, excepting average years of education of other members.

Table A2.2. Estimates from the bivariate probit model for the job equation (formal=1)

	Total	Men	Women
	Coefficient	Coefficient	Coefficient
Schooling years	0.0720**	0.0597**	0.0889**
	[0.0051]	[0.0063]	[0.0084]
Experience	0.0309**	0.0345**	0.0283**
	[0.0042]	[0.0057]	[0.0065]
Experience^2	-0.0006**	-0.0006**	-0.0006**
	[0.0001]	[0.0001]	[0.0002]
Tenure (months)	0.0089**	0.0070**	0.0117**
	[0.0005]	[0.0007]	[0.0009]
Tenure^2	-0.0000**	-0.0000**	-0.0000**
	[0.0000]	[0.0000]	[0.0000]
Women	0.1124**		
	[0.0392]		
Married	0.1578**	0.1240*	-0.0284
	[0.0490]	[0.0502]	[0.0448]
Women Married	-0.2146**		
	[0.0625]		
Household head	0.0123	0.0002	-0.0674
	[0.0472]	[0.0475]	[0.0534]
Women Household head	-0.1344*		
	[0.0668]		
Industry	-0.1677+	-0.1422	-0.1744
	[0.0895]	[0.1025]	[0.1754]
Construction	-0.4423**	-0.4828**	-0.3387
	[0.0961]	[0.1072]	[0.2387]
Sales, Hotels and Restaurants	-0.0913	-0.0659	-0.0949
	[0.0893]	[0.1028]	[0.1739]
Transportation	-0.156	-0.1542	-0.1814
	[0.0965]	[0.1112]	[0.1860]
Financial Intermediation	0.1458	0.1435	0.1448
	[0.0964]	[0.1131]	[0.1829]
Social Services	-0.3247**	-0.3393**	-0.3206+
	[0.0926]	[0.1132]	[0.1744]
Firm Size Small	1.0304**	1.0576**	1.0007**
	[0.0319]	[0.0418]	[0.0499]
Firm Size Medium	1.4108**	1.4002**	1.4456**
	[0.0556]	[0.0698]	[0.0925]
Firm Size Large	1.6865**	1.7470**	1.6118**
	[0.0331]	[0.0449]	[0.0506]
Metro. Area U. Rate	-0.0285**	-0.0190*	-0.0413**
	[0.0069]	[0.0091]	[0.0105]

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

Table A2.2. Continued

	Total	Men	Women
	Coefficient	Coefficient	Coefficient
Average years of education other members	0.0301**	0.0360**	0.0243**
	[0.0046]	[0.0060]	[0.0071]
Number of kids age 0	-0.0916*	-0.1150*	-0.0344
	[0.0439]	[0.0528]	[0.0771]
Number of kids age 1	-0.0803+	-0.0920+	-0.0666
	[0.0432]	[0.0558]	[0.0680]
Number of kids age 2	-0.0573	-0.0053	-0.1549*
	[0.0429]	[0.0553]	[0.0692]
Number of kids age 3	-0.0761+	-0.0594	-0.0871
	[0.0449]	[0.0562]	[0.0732]
Number of kids age 4	-0.0886*	-0.0965+	-0.086
	[0.0441]	[0.0564]	[0.0694]
Number of kids age 5	-0.0534	-0.0158	-0.0797
	[0.0453]	[0.0580]	[0.0724]
Number of kids age 6	-0.0973*	-0.0473	-0.1701*
	[0.0463]	[0.0587]	[0.0735]
Number of kids age 7	-0.1567**	-0.1384*	-0.1752**
	[0.0435]	[0.0579]	[0.0658]
Number of kids age 8	-0.1525**	-0.0986	-0.2039**
	[0.0446]	[0.0604]	[0.0671]
Constant	-1.5313**	-1.5984**	-1.3881**
	[0.1395]	[0.1748]	[0.2426]
Observations	15104	8629	6475

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

Table A2.3. Estimates from the univariate probit overeducation model for men and women

	Men		Women	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Formal	-0.1321*	-0.0186*	-0.1457*	-0.0272*
	[0.0518]	[0.0073]	[0.0579]	[0.0109]
Schooling years	0.2700**	0.0381**	0.2166**	0.0404**
	[0.0078]	[0.0017]	[0.0087]	[0.0018]
Experience (years)	-0.001	-0.0015	0.0190*	0.0011*
	[0.0081]	[0.0004]	[0.0082]	[0.0005]
Experience^2	-0.0002	-	-0.0004+	-
	[0.0002]	-	[0.0002]	-
Tenure (months)	-0.0003	0.0000	-0.0009	-0.0002
	[0.0008]	[0.0001]	[0.0009]	[0.0001]
Tenure^2	0.0000	-	0.0000	-
	[0.0000]	-	[0.0000]	-
Married	0.0111	0.0016	-0.0505	-0.0094
	[0.0524]	[0.0074]	[0.0458]	[0.0086]
Household head	-0.0283	-0.004	0.0744	0.0139
	[0.0525]	[0.0074]	[0.0532]	[0.0099]
Industry	0.1432	0.0202	-0.04	-0.0075
	[0.1068]	[0.0150]	[0.1507]	[0.0281]
Construction	0.6339**	0.0894**	-0.1767	-0.033
	[0.1207]	[0.0171]	[0.2187]	[0.0408]
Sales, Hotels, Restaurants	-0.1376	-0.0194	-0.0354	-0.0066
	[0.1082]	[0.0153]	[0.1490]	[0.0278]
Transportation	-0.4494**	-0.0634**	-0.1926	-0.0359
	[0.1207]	[0.0173]	[0.1590]	[0.0297]
Financial Intermediation	-0.5263**	-0.0742**	-0.2867+	-0.0535+
	[0.1167]	[0.0167]	[0.1533]	[0.0286]
Social Services	-0.7589**	-0.1070**	-0.3402*	-0.0635*
	[0.1214]	[0.0173]	[0.1487]	[0.0277]
Firm Size Small	-0.0945	-0.0133	0.0594	0.0111
	[0.0584]	[0.0082]	[0.0640]	[0.0120]
Firm Size Medium	-0.2305*	-0.0325*	-0.1645	-0.0307
	[0.0896]	[0.0126]	[0.1012]	[0.0189]
Firm Size Large	-0.1092+	-0.0154+	0.1421*	0.0265*
	[0.0573]	[0.0081]	[0.0609]	[0.0114]
Metro. Area U. Rate	0.0049	0.0007	0.0082	0.0015
	[0.0108]	[0.0015]	[0.0114]	[0.0021]
Constant	-3.9013**	-	-3.8544**	-
	[0.2000]	-	[0.2373]	-
Observations	8890		6785	
Log pseudolikelihood	-2800.57		-2384.24	

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

Chapter 3. Double Penalty in Returns to Education: Informality and Educational Mismatch in the Colombian Labor market

3.1 Introduction

A number of explanations have been offered to explain why some earning-relevant characteristics, for example, education, are better rewarded in the formal sector than in the informal sector. An important bulk of these explanations is based on a segmented view of the labor market. For instance, the presence of restrictive labor market institutions and strict regulation of entry into the formal sector could pose a possible cause, so that some workers that do not have access to the formal sector are forced to accept informal sector jobs characterized by inferior earnings (see Fields, 1975). However, none of the former studies have considered one aspect which can affect the wage gap between formal and informal workers, that is, the way workers match their acquire education to the one required to perform their job. Independently of the method used to measure skill mismatches, a number of studies that estimated the effects of overeducation on earnings for developed and developing countries found that overeducated workers tend to earn higher returns to their years of schooling than co-workers who are not overeducated, but lower returns than workers with similar education who work in jobs that require the level of education that they possess[1]. In the previous chapter, it was found that after controlling for other characteristics and correcting for endogeneity, informal salary workers are more likely to be overeducated than formal workers. Thus it is possible that the formal-informal wage gap is driven, at least in part, by a less satisfactory matching of education-occupation in the informal sector and by the penalization in terms of wages that is derived from this mismatch. Actually the aim of this chapter is to re-examine the wage gap between formal and informal workers taking into consideration that education-occupation mismatch is present in both sectors, using the case study of Colombia.

This study contributes to the literature on informality and education-occupation mismatch by gauging if the return to years of required education, years of surplus education and years of deficit education differ across formal

[1] For an extensive review of overeducation in developed countries see McGuinness (2006) and for a recent survey on overeducation see Leuven and Oosterbeek (2011).

and informal sectors. If they do differ and if salaried informal workers are more penalized in terms of wages in the presence of educational mismatches than their formal counterparts, then we can infer that part of the formal-informal wage gap might be originated in such a difference. A similar approach is adopted in Chiswick and Miller (2008) in their analysis of the difference in returns to education between native and foreigners in United States. These authors find that the lower payoff to schooling for foreign-born workers is due to under education (linked with positive self-selection in immigration among immigrants with low levels of schooling) rather than to overeducation (related to the less-than-perfect international transferability of human capital). Under the same line, Ren and Miller (2012) also use the over-under educated framework for analyzing the difference in the returns to schooling between men and women in China. As far as we know, the idea of distinguishing the difference in the returns from correct, over and deficit years of education for formal and informal workers is a novel contribution, as there is no previous study that considered this difference before in all analyses of which we know about informality[2].

The empirical analysis consists of examining the returns to education taking into consideration the existence of educational mismatches in the formal and informal sector. For this purpose we first estimate the standard Duncan and Hoffman (1981) specification (so called ORU wage equation) at the mean, using ordinary least square (OLS), and controlling for a rich set of observable individual and firm characteristics. Then, we examine if the returns to education for each of the education-occupation mismatch are not uniform along the wage distribution by using quantile regression estimation. In both cases the endogeneity sector choice is addressed. Finally we implement a decomposition developed by Chiswick & Miller (2008), which allows disentangling the effect of educational mismatch in the difference in the returns to education. It does this by distinguishing the contribution of the returns to years of overeducation, required education and under education to the difference in the returns to education in the conventional (Mincer) human capital equation.

[2] See, for example, Magnac (1991), Nuñez (2002), Maloney and Nuñez (2004), Flórez (2002), Kugler and Kugler (2009) and Mondragón-Vélez, Peña, and Willis (2010) for Colombia; Gindling (1991) for Costa Rica; Pradhan and van Soest (1995) for Bolivia; Amuedo-Dorantes (2004) for Chile; Pratap and Quintin (2006) for Argentina; Tansel (2000) for Turkey; Marcouiller, Ruiz de Castilla, and Woodruff (1997) and Gong and van Soest (2002) for Mexico; Botelho and Ponczek (2011) for Brazil; Badaoui, Strobl, and Walsh (2008) for South Africa.

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

The structure of the chapter is organized as follows. The next section gives a description of the data and some selected descriptives, while the empirical approach is presented in section 3.3. Section 3.4 summarizes the results regarding the estimates of the empirical models for the Mincer and ORU wage equations. Section 3.5 presents the results from the decomposition analysis, in section 3.6 some gender issues are examined and, finally, section 3.7 concludes.

3.2 Data and descriptive statistics

In this study, a sample of 34626 working individuals was drawn from the 2010 CHS[3]. The analysis was restricted to salary workers that were not carrying formal studies aged between 15 and 60 years and who report working more than 16 hours per week. We do not include self-employed and employers workers in the analysis because their source of income is a combination of labor and physical capital and therefore may not be compared with earnings of other employees. Apart from this, self-employed workers' earnings would be expected to have a greater measurement error. Also, while comparing self-employed informal workers to their formal counterparts may be of interest, it has been shown in previews studies that self-employed in the informal sector corresponds more with a voluntary entry, while informal salaried work may correspond more closely to the standard queuing view, especially for younger workers (Perry et al., 2007; Bosh and Maloney, 2010). Excluding self-employed resulted in dropping 16941 individuals. We also exclude public employees from the sample since by nature they belong to the formal sector and their wages might reflect institutional arrangements. After excluding observations with missing values or inconsistencies for the selected regressors, over 13797 individuals remained in our sample.

As in chapter 2, we classify workers as formal or informal according to whether they are covered by the social security system or no and for the purpose of measuring the incidence of the education-occupation mismatch we also define required education using the statistical method in its mean version. Under the statistical method required education is define as the mean level of schooling for each occupation. Individuals are classified as overeducated (undereducated) for a particular occupation if their level of education is higher (lower) than the required education. In the mean measure a worker is overeducated or under-educated if their completed level of schooling deviates by one standard deviation from the mean in their occupation.[4] As starting point formal and informal workers are pooled together for obtaining the correct level of education. Later, as a sensitive analysis, formal and informal workers are treated separately for calculating the years of correct education.

[3] The national statistic department, DANE, through its web page made available the data for 2010 after completing the first study presented in chapter 2. However, the descriptive statistics that will be presented as well as the results are not strongly affected by the year of the database.

[4] For purpose of brevity we only included the results obtained with the mean, as with the mode the results are not significantly different. An appendix with the full set of results is available on request.

Regarding earnings, we have combined information from gross monthly income and worked hours in order to obtain gross hourly wages.

Table 3.1 contains mean hourly wages by job type and educational mismatch. As it can be seen informal workers, in average, earn less than formal workers, informal workers earn 78 per cent less than what formal workers earn for the total sample. This large wage differential found here is in line with the findings of several other studies for other countries, and so far has been the centerpiece of the empirical analysis in the past. If formal and informal workers are classified by educational mismatch the wage gap is not the same across the different categories. For instance, overeducated formal workers earn 90 per cent more than informal overeducated workers, while undereducated formal workers earn 40 per cent more than their informal peers. The formal – informal wage gap is also higher for the overeducated than for workers correctly matched in terms of education.

Table 3.1 also presents the formal-informal wage gap at different quantiles. As it can be seen the wage gap is not homogeneous along the wage distribution and across the different education-occupation mismatches. The first thing to be noticed is that the hourly wage at the lower quantile for correct and overeducated formal workers are both equal to the minimum wage[5], while an undereducated formal worker perceives a wage slightly lower[6]. This finding conforms to the notion that the minimum wage is binding in the formal sector. The formal-informal wage gap among the least skilled, measured by the lower quantile of the wage distribution, is considerably lower for overeducated workers compare to correct and undereducated workers. This could be indicating that a formal worker in the lower part of the distribution and regardless of his education will be rewarded with a wage similar to the minimum wage, while informal wages are determined freely. This possibility to set wages freely allows informal sector to pay a considerably lower wage to correct and undereducated workers, while somehow rewarding overeducated workers. In contrast, at the middle and, particularly, at the upper part of the distribution the formal-informal wage gap is substantially higher for overeducated workers compare to correct and undereducated workers. Thus,

[5] The monthly minimum wage in Colombia in 2010 was 515,000 pesos, equivalent to 2503.47 pesos per hour (this value is obtain by first dividing the monthly minimum wage by 4.3 to obtain weekly wage which in turn is divided by 48 weekly hours of work to reach hourly wage).

[6] A close inspection of the data shows that on average undereducated workers at the lower part of the distribution earn a wage equal to the minimum monthly wage, however as some undereducated workers reported working more than 48 hours the wage observed at the lower quantile is slightly less than the computed minimum hourly wage.

this simple preliminary evidence, at the mean and at different quantiles, indicates that educational mismatch may be a key aspect in order to get a better understanding of the formal – informal wage gap.

Table 3.2 presents the distribution of years of overeducation, required education and under education by years of actual education. The incidence of correctly educated workers is similar across the two types of employment, around 76% for formal and 74% for informal. Formal workers seem more likely to be overeducated, 17%, compare to their informal counterparts, 14%, while informal workers seem more likely to be undereducated, 15%, than formal workers, 7%. It is interesting that conditional upon a particular category of actual years of education, there are only negligible differences in the degree of under education between formal and informal workers. The difference in the degree of overeducation between formal and informal workers is more evident. For example, in almost all categories of actual years of education informal workers are more likely to have more than three years of surplus education than formal workers. From Table 3.2 it is deduced that formal workers seems more likely to be overeducated than informal workers, while informal workers are more likely to be undereducated than their formal counterparts. However, these differentials in the incidence of over- and under education may just be caused by a composition effect, in other words, formal workers are more educated whereas informal worker are less educated. The preview chapter showed that a sorting effect drives the gap in the raw propensities, and, that when comparing formal and informal workers with similar individual and firm characteristics, those in the former group have a lower propensity to be overeducated.

Table 3.3 presents some basic summary statistics concerning the distribution of the observed workers' and firms' characteristics that may be driving the wage differentials[7]. It shows information for the entire sample of workers, and distinguishing between those working in the formal and in the informal sectors. Formal workers in our sample are more likely to have higher education or more (44 per cent), whereas informal workers are more likely to have basic secondary and secondary (22 per cent and 36 per cent respectively). There is not significant difference in age and experience between workers in both groups. In contrast, there are some notable differences in the average tenure between sectors; formal workers tend to accumulate much more tenure than informal workers, suggesting higher stability of employment for formal

[7] Comparing Table 3.3 to Table 2.1 of Chapter 2 it can be seen that the descriptive statistics do not differ significantly between 2008 and 2010.

workers. As a matter of fact, 95 per cent of formal workers had signed a contract, and 65 per cent of them of a permanent type, in contrast with only 18 per cent of informal workers who have a contract, and only 10 per cent having a permanent one. On the other hand, as can be seen, the percentage of female workers in the formal sector is slightly higher than in the informal. This may be due to the fact that our sample excludes self-employed individuals and unpaid family workers. A much larger proportion of the workforce in the formal sector is married. In terms of the occupational structure, informal workers are more likely to be found in unskilled manufacturing and agricultural occupations (43 per cent). Formal are also more likely to be found in unskilled manufacturing and agricultural occupations, but at a lower rate (25 per cent), followed by administrative staff (24 per cent). There is little difference in the average hours of work in the two sectors. As for the firm size, as expected and given the high relationship between informality and size, firms with less than 3 regular employees are substantially more likely to be part of the informal sector. In contrast, larger firms employ much of the formal-sector labor force with a workforce greater than one hundred.

3.3 Empirical strategy

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

$$W_{ij} = \alpha_j S_{ij} + \beta_j X_{ij} + \varepsilon_{ij} \qquad (3.1)$$

where W_{ij} denotes the log of the hourly wage of the individual *i* in sector *j*, formal (*F*) or informal (*I*), S_{ij} the years of acquire education, X_{ij} denotes the set of other characteristics (for example experience, tenure, gender) that affect the wage of this individual; α_j is the return to years of acquire education and β_j is a vector of prices or returns associated with other characteristics that affect wages. Finally, ε_{ij} is the error term for individual *i* in sector *j*.

The typical specification adopted to estimate the effect on earnings of education – occupation mismatch is based also on the Mincerian wage equation. However, the general educational mismatch specification varies slightly in that the variable of acquired years of schooling is decomposed into

three variables: required, surplus and deficit education, following Duncan and Hoffman (1981) formulation. Overeducation is the amount of years of schooling a worker has acquired in excess of the required education needed to perform his job. Under education entails the opposite. Under this framework wages are a function of over, required and deficit years of education (so-called ORU wage equation). That is:

$$W_{ij} = \alpha_{rj} S_{ij}^{r} + \alpha_{oj} S_{ij}^{o} + \alpha_{uj} S_{ij}^{u} + \beta_j X_{ij} + \nu_{ij} \qquad (3.2)$$

where S^r is years of required education, S^o is years of surplus education above the required level and S^u is years of deficit schooling below the required level[8]. Then, under this wage equation the returns from additional education are α_{rj} for required years, α_{oj} for surplus years, and α_{uj} for deficit years of education. Notice that instead of imposing the same return in the two sectors, we allow them to differ for workers in each sector *j*, formal or informal.

Next we want to analyze the returns to education and the effects of occupation-education mismatch on the entire wage distribution for formal and informal workers, by using linear quantile regression estimates. By estimating linear quantile regressions we are able to examine the heterogeneous effect of education at different points in the wage distribution. Moreover, quantile regression estimates are robust to the outliers of the dependent variable and they are also more efficient than the OLS under non-normality of the error terms. For any worker *i* in sector *j* we can write the $\boldsymbol{\tau}^{\text{th}}$ quantile of the hourly wage distribution conditional on actual years of education (S_{ij}) and other characteristics (X_{ij}) as:

$$F_{Wij}^{-1}\left(\tau \middle| S_{ij}, X_{ij}\right) = S_{ij}\alpha_j(\tau) + X_{ij}\beta_j(\tau), \forall\, \tau \in [0,1] \qquad (3.3)$$

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

[8] Years of acquire education equals years of required education plus years of surplus education minus years of deficit education ($S = S^r + S^o - S^u$).

(S^o_{ij}), years of deficit education (S^u_{ij}), and other characteristics (X_{ij}) as:

$$F^{-1}_{Wij}(\tau|S_{ij}, X_{ij}) = S^r_{ij}\alpha_{rj}(\tau) + S^o_{ij}\alpha_{oj}(\tau) + S^u_{ij}\alpha_{uj}(\tau) + X_{ij}\beta_j(\tau), \forall\, \tau\ [0,1] \qquad (3.4)$$

The specifications formulated so far (eqs. 3.1 to 3.4) neglect the existence of non-observable characteristics that could simultaneously affect wages and the sector in which the individuals are currently working. This will cause to obtain not only biased, but also inconsistent coefficients of the return to education. To account for this concern, we implement the conventional approach of including a selection correction in the wage regressions for each sector. This entails a two-stage estimation process. In a first stage a reduced-form probit model of the formal vs. informal decision is estimated, and a sample selection correction term is obtained. In stage two, the correction term is incorporated into conventional Mincerian semi-log earnings functions for the formally employed and informally employed (see, for example, Gong and van Soest, 2002; Günther and Launov, 2012).

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

$$E_i^* = \gamma Z_i + \mu_i \qquad (3.5)$$

where E_i^* is a latent variable that determines the sector j (= formal, informal) in which individual i is employed, Z_i is a vector of observed individual characteristics included in X_i in the wage equations plus some other variable(s) likely to affect the propensity to be employed in the formal or informal sector, and μ_i is the error term.

The observed binary variable E_i is related to the latent variable E_i^* as follows:

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

Estimates of returns based on the wage eq. (3.1) to eq. (3.4), leaving aside the selection eq. (3.5), are biased and inconsistent if the error term of the selection equation and the error terms of the wage equations are correlated, for example $cov[\mu_i, \varepsilon_{ij}] = \rho_j \neq 0$ for the mean Mincerian wage eq. (3.1).

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

While the methods for correcting sample selection for mean regression are well acknowledged, there are few known approaches to correct for selectivity bias in quantile regression models and there is little consensus regarding the most appropriate correction procedure. Buchinsky (1998) suggests an approach to approximate the selection term by a power series expansion of the inverse of the Mill's ratio and is the most common approach used so far for correcting selectivity in quantile regression models (Garcia, Hernández, and López-Nicolás, 2001; de la Rica, Dolado, and Llorens, 2008; Albrecht, van Vuuren, and Vroman, 2009). We thus follow Buchinsky (1998) procedure for correcting the potential selection bias in the estimation of wage equations that may result from the selection of workers into formal or informal jobs.

3.4 Results

3.4.1 OLS results

Table 3.4 presents the coefficients obtained from estimating the Mincer wage equation (3.1) and the coefficients of estimating the ORU wage equation (3.2). Estimates were done separately for formal and informal workers. A simple specification for the two wage equations was used to account fully the effect of human capital variables. It includes as explanatory variables the number of years of education (actual years of education in the Mincerian wage equation and years of education decomposed into surplus, required and deficit in the ORU wage equation), the years of experience and its square, the months of tenure with the current firm and its square, and the gender of the individual. The results of this simple specification are presented in the first column of each estimated wage equation. However, as it has been shown in the descriptive analysis, formal and informal workers differ significantly in firm and individual characteristics, beside those related to human capital. For instance, given that firms tend to be larger in the formal sector and larger firm

pay more, formal workers could obtain a higher return to their education just because they are more prone to work in large firms while informal workers are more likely to work in small firms. Thus to ensure that the comparison of the returns to education across the two sectors is done for observably similar workers, a more comprehensive specification that includes additional controls was used for the two wage equations. Besides, including additional individual and job characteristics also allow us to disentangle to what extend these observable characteristics explain the average wage differentials across formal and informal workers. Those controls include dummy variables for marital status, head of household, occupation, contract signed, size of the firm, industry sector, hours worked and a dummy variable indicating the metropolitan area. The results of this more comprehensive specification are shown in the second column of each estimated wage equation in Table 3.4.

We start by describing the results of the Mincerian wage equation for the simple specification (columns labeled 1). The results show that education is better rewarded in the formal sector than in the informal sector, since each additional year of schooling increases hourly wages by 10.08 per cent for formal workers, which is around double that for the informal workers, 5.43 per cent. As expected, once additional controls are accounted for (columns labeled 2) the return to schooling estimated for both sectors is lower. Each additional year of schooling increased hourly wage by 9.00 per cent for formal workers and by 4.19 per cent for informal workers. Nevertheless, the finding that formal workers have a higher return to their education than informal workers still holds.

Considering the existence of educational mismatches gives an interesting picture of the difference in the returns to schooling across the two sectors. Table 3.4 also presents the returns associated with schooling when educational mismatches are present –the ORU wage equation (3.2). Consistent with previous literature i) the returns to surplus schooling are lower than the returns to required schooling, ii) a year of deficit schooling carries a wage penalty for both sectors, and iii) the returns on required education are higher than that on actual or attained education in the Mincer equation. As it can be seen, the returns to required and to surplus schooling are higher in the formal sector than in the informal. Results from the specification that does not include the full set of controls indicate that one additional year of required education raises hourly wages by 13.23 per cent in the formal sector and by 7.63 per cent in the informal. Years of surplus education are associated with an earning increase of 9.31 per cent for formal workers and 4.16 per cent for

informal workers. Noteworthy is that the penalty of deficit schooling is not very dissimilar across the two sectors, 3.36 per cent for formal workers and 4.68 per cent for informal workers. As for the results when additional controls are introduced in the estimation of the ORU wage equation, it can be observed that the returns to required and surplus schooling diminish but only slightly, whereas the decrease in the estimate of the penalty of deficit schooling is more intense for informal workers. In any case, regardless of the inclusion or not of additional controls, results confirm that the returns to required and surplus education for formal workers are significantly higher than those for informal workers.

As the years of required education for each occupation can be calculated jointly for formal and informal workers or separately, it is possible that the results shown above may vary depending on how the reference group for determining the required years of education is selected. On the other hand, it could also be said, that formal jobs are characterized by a better selection process than informal jobs, so in order to obtain a more accurate measure of the education required to perform a job, it might be more appropriate to use only information about formal workers. Thus, the ORU equation was also estimated when the years of required education were calculated individually for formal and informal workers and when formal workers were the only reference group for calculating the years of education required to perform a job. Table 3.5 presents these results. As can be seen the returns to over, under and required years of education are not sensitive to the selection of the reference group for calculating the years of required years of education.

To sum up, formal workers have higher returns to their years of education than informal workers, and this is so in the presence of educational mismatch. Moreover, overeducated informal workers are double penalized, since in addition to the lower return to years of required education for the fact of being in the informal sector, they face a second penalty associated with the lower returns they obtain because of the discrepancies between workers' actual years of education and the level of education required for performing their job, that is considerably larger than that for their formal counterparts.

3.4.2 Quantile results

The OLS results provide the return estimates at the mean of the wage distribution, which may be hiding important differences in the return estimates

at different points of the wage distribution. Table 3.6 presents the quantile regressions results obtained from estimating the Mincerian wage equation - eq. (3.3) - in the upper panel and the ORU wage equation - eq. (3.4) – in the lower panel. Both equations were estimated using all set of controls (dummy variables for marital status, head of household, occupation, contract signed, size of the firm, sector industry, hours worked and metropolitan area).[9] To facilitate the comparison of results at the different quantiles with those at the average, results of the OLS estimates are reproduced in the first group of columns in Table 3.6. The results reveal that schooling is not uniformly rewarded in the labor market along the wage distribution. More specifically, the return to actual education (upper panel of Table 3.6) increases along the wage distribution for formal workers, while a comparable pattern is not observable for informal workers. Thus, education may contribute to generate important wage differentials among formal and informal workers. Under the observed wage structures, more years of schooling would make the distribution of formal wages more disperse, but informal workers wages´ dispersion would not experience any significant increase. Interestingly, the difference in the return to actual education for formal and informal workers in the 25th quantile is minimal (4.61 per cent versus 3.23 per cent), while at the 75th quantile the return to actual education for formal workers is around three times higher than that for informal workers (9.99 per cent versus 3.39 per cent). That the returns to education for formal workers in the 25th quantile are very similar to those of informal workers counterparts can be the result of the existence of a minimum wage, binding only for the formal sector, which could be imposing an important distortion to the returns to education to formal sector workers at this part of the distribution.

The results obtained for the ORU specification in eq. 3.4 (bottom panel of Table 3.6) show that the behavior of the returns to required education resembles that of actual education: they increase substantially along the wage distribution for formal workers, but only experience a moderate change for informal workers. Remarkably, results also suggest that the returns to surplus education behave similarly, increasing along the wage distribution for formal workers and remaining almost constant across the different quantiles for informal workers. In turn, the pattern of the penalty associated to deficit education is different for formal and informal workers, although the order of magnitude of the difference in this case is much lower than for

[9] Similar results were obtained with the simple specification that does not include the additional set of controls. They are available from the authors.

required and surplus education. A clearer picture of these patterns is obtained by plotting the estimated returns at each percentile for formal and informal workers as in Figure 3.1 As it can be seen, returns to education are not homogenous along the wage distribution and this heterogeneous behavior is very different for formal and informal workers.

A more detailed inspection of the lower panel of Table 3.5 reveals additional key information. For instance, differences in the educational returns between formal and informal workers with the same educational-occupational mismatching are present at the 25th quantile, although less sizeable than the differences presented in the 75th quantile. Formal workers that possess the education required to do their job have a higher return to their education, slightly higher in the lowest quantile and more than double in the upper. An overeducated formal worker in the lower part of the distribution obtain a return of his years of surplus education similar to the return obtained by an informal worker for the years of education required to perform his job, 4.46 per cent and 4.73 per cent respectively. Meanwhile the returns to surplus education for formal workers at 75th quantile of the distribution are larger than the returns to required education for informal workers, 9.63 per cent and 5.65 per cent correspondingly.

Summing up, the results from the quantile regression lead to the conclusion that formal workers are able to obtain a higher reward for their education even in the presence of educational mismatch, and this is so along all the wage distribution. Furthermore, the returns to surplus education increase considerably for formal workers along the wage distribution suggesting that this type of jobs represents better employment opportunities for overeducated workers. This probably reflects the fact that formal workers may take advantage of the higher productivity[10] that is present in formal jobs, which may boost the returns to education. Meanwhile, informal workers receive a lower remuneration to their education compared to the one obtained by their formal peers. This difference in returns to education between formal and informal workers is even more accentuated in the upper part of the distribution. More importantly, informal overeducated workers do not face higher returns once they move up the wage distribution, implying that informal jobs may constraint the use of education and its returns.

[10] 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 (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.

3.4.3 Sample selection

Our estimates of the wage equations, when taking into account that unobservable variables might influence both wages and the choice of formal/informal employment, are summarized in Table 3.7 for the estimates at the mean. These results correspond to estimates of the wage equations augmented by a selection correction term for each sector, using the presence of children in the household and the average number of years of schooling of other household members as additional determinant of the assignment into the formal or informal sector. The reason for choosing these selection variables is motivated by the fact that they should contain household-specific characteristics that influence an individual's choice regarding formal or informal employment, but at the same time have no direct impact on the earning potentials of individuals (Günther and Launov, 2012 use similar variables as exclusions restrictions). As it can be seen, once the selectivity is corrected the returns to schooling remains higher for formal workers in the two wage equations estimated (Mincer and ORU). It is important to note that the selection term (*Mills lambda*) is positive and statistically significant for formal workers. This result can be interpreted as follows: a worker that has a higher probability of working in the informal sector, due to his observable characteristics, could end up working in the formal sector thanks to unobservable factors (for example, job-search networks or ability) and gets a higher return to his education (Tannuri-Pianto, Pianto, and Arias, 2004 find a similar result for Bolivia). In the case of informal workers the selection term is insignificantly different from zero. This implies that there is no correlation between the error terms of the selection equation in (3.5) and that of the wage equation for informal workers, and thus that the estimates given in Table 3.4 for informal workers seem to be unbiased.

We also re-estimate the quantile regressions of eq. (3.3) and eq. (3.4) introducing the inverse of the Mills's ratio and its square, following the Buchinsky (1998) procedure for correcting for selection bias. The results are presented in Table 3.8. It can be observed that the pattern of estimated returns and differences between formal and informal workers reported and discussed in the previous section do not vary significantly when selection is accounted for.

All in all, from these results we can assert that the major conclusion on the higher penalty associated to educational mismatch for informal workers

remains when controlling for the correlation between the error terms in the selection and the wage equations.

3.5 Decomposing the formal–informal gap in returns to education

We move the analysis a step further and implement a decomposition developed by Chiswick & Miller (2008), which allows disentangling the effect of educational mismatch in the difference in the returns to education. It does this by distinguishing the contribution of the returns to years of overeducation, required education and under education to the difference in the returns to education in the conventional (Mincer) human capital equation. This approach was adopted in Chiswick and Miller (2008) to analyze the difference in returns to education between native and foreigners in United States. These authors find that the lower payoff to schooling for foreign-born workers is due to under education (linked with positive self-selection in immigration among immigrants with low levels of schooling) rather than to overeducation (related to the less-than-perfect international transferability of human capital). Under the same line, Ren and Miller (2012) also use the over-under education framework for analyzing the difference in the returns to schooling between men and women in China. As far as we know, this decomposition analysis for understanding the difference in returns to schooling for formal and informal workers is a novel contribution, as there is no previous study that has done it in all analyses of which we know about informality.

3.5.1 Decomposition framework

For simplicity purposes, we propose a slightly modified exposition of the original decomposition proposed by Chiswick and Miller (2008). The decomposition involves the construction of hypothetical workers for each level (years) of education $l=1,\ldots, L$[11]. In the first step of the decomposition, each hypothetical informal worker at each educational level (l) is assigned the mean years of over, require and under education of informal workers calculated at the educational level to which he corresponds. Furthermore, in order to standardize for variation in other characteristics not related to education, all hypothetical workers are given the mean levels of all the

[11] 19 levels of education were constructed.

characteristics included in X_{ij}. Then using the coefficients from the estimated ORU equation (2) of informal workers, a wage for each hypothetical worker at each educational level is predicted as follows:

$$(\widehat{W}_1)_l = \hat{\alpha}_{rI}(\overline{S_I^r})_l + \hat{\alpha}_{oI}(\overline{S_I^o})_l + \hat{\alpha}_{uI}(\overline{S_I^u})_l + \hat{\beta}_I\overline{X_I} \text{ , for } l=1,\ldots, L \quad (3.6)$$

Then each of these predictions is regressed on the level of education. In this supplementary simple regression, each observation is weighted by the number of informal workers with the particular level of education (θ_{Il}). That is:

$$(\widehat{W}_1)_l\theta_{Il} = \alpha_1 S_{Il}\theta_{Il} + \mu_{Il}\theta_{Il} \text{ , where } \theta_{Il} \text{ are the weights for } l=1,\ldots, L \quad (3.7)$$

In this weighted simple regression, $\hat{\alpha}_1$is an estimate of the return to actual education for informal workers, similar to the one that is obtained from eq. (3.1) using the individual-level data for informal workers.

In the second step, the coefficients from the estimated ORU equation (3.2) of informal workers ($\hat{\alpha}_{rI}, \hat{\alpha}_{oI,}\hat{\alpha}_{uI}$) are replaced for the coefficients estimated from the sample of formal workers. That is:

$$(\widehat{W}_2)_l = \hat{\alpha}_{rF}(\overline{S_I^r})_l + \hat{\alpha}_{oF}(\overline{S_I^o})_l + \hat{\alpha}_{uF}(\overline{S_I^u})_l + \hat{\beta}_I\overline{X_I}, \text{ for } l=1,\ldots, L \quad (3.8)$$

Then each of these predictions is regressed on the level of education. In this supplementary simple regression, each observation is weighted by the number of informal workers with the particular level of education (θ_{Il}). That is:

$$(\widehat{W}_2)_l\theta_{Il} = \alpha_2 S_{Il}\theta_{Il} + \mu_{Il}\theta_{Il} \text{ , where } (\theta_{Il}) \text{ are the weights for } l=1,\ldots, L \quad (3.9)$$

In this second supplementary regression, $\hat{\alpha}_2$ is the return to education for informal workers under the assumption that the returns to over, under and required years of education are the same for informal and formal workers. Comparison of this return with that obtain using the prediction of eq. (3.7) reveals the contribution to the differences in the estimated effect of the ORU variables for formal and informal workers to the conventional estimate of the return to education for informal workers.

In the third step, the ORU variables for informal workers are replaced using the sample average, conditional upon a particular level of education, for formal workers. That is:

$$(\widehat{W}_3)_l = \hat{\alpha}_{rF}(\overline{S_F^r})_l + \hat{\alpha}_{oF}(\overline{S_F^o})_l + \hat{\alpha}_{uF}(\overline{S_F^u})_l + \hat{\beta}_I\overline{X_I} \text{ , for } l=1,\ldots, L \qquad (3.10)$$

Then each of these predictions is regressed on the level of education. In this supplementary simple regression, each observation is weighted by the number of informal workers with the particular level of education (θ_{Il}). That is:

$$(\widehat{W}_3)_l\theta_{Il} = \alpha_3 S_{Il}\theta_{Il} + \mu_{Il}\theta_{Il}, \text{ where } \theta_{Il} \text{ are the weights for } l=1,\ldots, L \qquad (3.11)$$

In this third supplementary regression, $\hat{\alpha}_3$is the estimate of the returns to education for informal workers under the twin assumption that the returns to the ORU variables for informal workers are the same as for formal workers and the mean values of these variables for informal are the same as for formal. This simulated return to education can be compared to that obtained in the previous step to assess the incremental contribution of differences in the values of the ORU variables for informal and formal workers to the return to education obtained by informal workers.

The final step in the decomposition involves using the number of formal workers at each level of education for the weighting variable in the supplementary weighted simple regression depicted in equation (3.12):

$$(\widehat{W}_4)_l\theta_{Fl} = \alpha_4 S_{Il}\theta_{Fl} + \mu_{Il}\theta_{Fl}, \text{ where } \theta_{Fl} \text{ are weights for for } l=1,\ldots, L \qquad (3.12)$$

Following this change, the $\widehat{\alpha}_4$ obtained from this simple regression will be the estimate of the return to education for formal workers.

The set of substitutions outlined above progressively move us from the return to education for informal workers to the return to education for formal workers. This enables the roles of matched and mismatched education in the labor market on the return to education to be assessed.

3.5.2 Decomposition results

The results of the decomposition exercise are presented in Table 3.9, they were computed correcting for sample selection. As can be seen in panel

A, the returns to actual education computed following the procedure of the decomposition exercise are 9.06% for formal workers (eq. 3.12) and 4.44% for informal workers (eq. 3.4); these values are very closed to those obtained with the estimation of the Mincer eq. (3.1) using the individual-level data (9.07% and 4.13% respectively, see Table 3.7 with sample selection). Once the returns associated with overeducation, under education and required education for informal workers were replaced by the respective returns estimated for formal workers, the payoff to schooling for the informal was found to be 7.45%. Meaning that informal workers would have an increase of 3.01 percent points in their returns to actual education if the effect associated with overeducation, under education and required education were the same as those for formal workers. In other words, the difference in returns to overeducation, under education and required education between formal and informal workers explains 64% of the difference in the return to the actual education of both groups (0.65 =7.45-4.44/9.06-4.44).

Replacing the information on the distribution of the formal workers across overeducation, under education and required education, at comparable levels of schooling, for the informal workers does not results in a significant change in the payoff to schooling for informal workers, since the value computed is 7.55%. The reason why there is no change in the payoff for informal workers in this case is that, conditional on the level of education; there are not significant differences between the distribution across under education and required education, though there are some difference in the distribution of overeducation (See Table 3.2). Finally if the distribution of actual years of education for informal workers is replace by the distribution of actual years of education of formal workers then the payoff to schooling for informal workers will increase by 1.51 percent points (1.51=9.06-7.55). This last step of the decomposition (eq. 3.12) resulted in a payoff to schooling for informal workers equal to that of formal workers, 9.06%. Therefore, the difference in the distribution of actual years of education between formal and informal workers explains 32.68% of the difference in the return to actual education of formal and informal workers (32.68=9.06-7.55/9.06-4.44). The disproportionate representation of informal workers among the lower education categories and the disproportionate representation of formal workers among the highest education categories of actual education are driving this result.

Panel B, C, D and E of Table 3.9 presents the results where adjustments in the decomposition are made only for required education,

overeducation, under education and required and overeducation, respectively. It is clear that the returns associated with required and overeducation explain an important mass of the gap between the payoffs to schooling for formal and informal workers, whereas the contribution of the effects of under education is minor. Of the 3.01 percentage points increase in the returns to actual education for informal workers if the effects of overeducation, under education and required education were the same as those for formal workers, 1.72 percentage points are related to adjustments of the returns to required education (Panel B: 1.72=6.16-4.44), 1.39 percentage points are linked to the return to overeducation (Panel C: 1.39=5.83-4.44) and -0.09 percentage points are associated to returns to under education (Panel D: -0.09=4.35-4.44). Thus 1.3 percentage points are related to the effects of educational mismatches.

With the regression analysis of the previous sections it was clear that the returns to correctly matched education and overeducation were much higher for formal workers than for informal workers. The decomposition analysis revealed that both of these factors contributed in a large extent to the higher returns to schooling for formal workers than for informal workers. In contrast, under education is a minor element for understanding the formal-informal gap in the returns to schooling.

3.6 Gender

In this section we evaluate the sensitivity of the results presented so far if they are conducted separately for men and women. Table 3.10 display the estimates of the Mincer and ORU wage equations, with and without correcting for sample selection for men and women. The results by gender are largely corresponded to the results for the total sample and they do not vary considerably with the correction for sample selection, especially in the case of men. However some differences emerge and are worth to be mentioned. First it should be noted that the returns to surplus years of education for informal male workers are considerably lower compared to those of formal male workers. For women the difference in returns to surplus education between informal and formal workers are lower to those found for male workers, though they remain significant. The most remarkable difference is that whereas informal male workers are greater penalized for their years of deficit education than formal male workers, the opposite happens in the case of female workers, female formal workers faced a stronger penalization to their

years of deficit education compared to the penalization faced by female informal workers.

Table 3.11 shows the returns to education from the Mincer and the ORU equations for the whole wage distribution and correcting for sample selection. Again the general conclusions derived from the total sample remain true for men. Nevertheless, specifically in the case of women, there are some findings that merit some attention. In contrast to what has been found previously, the returns to years of surplus education are higher in the 25th quantile for informal female workers compared to formal female workers. Additionally returns to required years of education at this quantile are not different between formal and informal female workers. In the middle part of the wage distribution, formal female workers obtain a considerable higher return to their years of required education compared to informal female workers and the returns that they obtain from surplus years of education are slightly higher. The penalizations to deficit years of education are higher for formal than informal female workers until the 50th quantile. The fact that the returns to years of surplus are higher for informal female workers at the lower part of the wage distribution may be a sign that low wage informal jobs for over educated women may be a more advantageous situation. Maybe informal jobs at this part of the wage distribution allow women to obtain a better reward to their skills because of the flexibility of hours of work and place of work. However, in the upper part of the wage distribution the results for female behave similar to those for male workers and informal jobs display considerable lower returns for surplus and required years of education.

Regarding the decomposition analysis, the results for men and women (presented in Table 3.12) largely corresponds to those found for the total sample. Similarly, we find that educational mismatches explain an important part of the gap between the returns to schooling for formal and informal for both collectives.

3.7 Conclusions

There is now substantial body of literature addressing the wage gap between formal and informal workers for developing countries, theoretically and empirically. In empirical analyses wage equations are estimated for each group of workers, where one of the key factors is education (and its returns). There are papers that have gone beyond the difference in the mean, finding

that the wage gap is not stable along the wage distribution, estimating quantile regressions. Some works have questioned the existence of a wage gap (that is, market segmentation) given the endogeneity caused by unobservable characteristics of the individuals, such as skills. As far as we know there is no study that considered the fact that education-occupation mismatching is present in both formal and informal sector, and that this may be driving, at least in part, the formal/informal wage gap. In this chapter we have re-examined the wage gap between formal and informal workers taking into consideration that education-occupation mismatch is present in both sectors, using the case study of Colombia.

Results for Colombia show that formal workers have a higher return to their education, around double, compared with their informal counterparts. They also indicate that these returns vary along the wage distribution and that the pattern of variation along the distribution is not the same for formal and informal workers. But on the top of that, the main claim in this chapter is that important information to the analysis of the formal–informal wage gap is obtained by adding measures of educational mismatch. In particular, we showed that the returns to required education in the informal sector are not only lower, but the penalty that informal workers face due to educational mismatches in terms of wages are considerable higher than the one faced by their formal counterparts. Therefore, we can conclude that there is a second penalty associated with educational mismatches that puts informal workers at a greater disadvantage compare to formal workers.

The decomposition analysis revealed that both of these factors contributed in a larger extent to the higher returns to schooling for formal workers than for informal workers. However, under education is a minor element for understanding the formal-informal gap in the returns to schooling.

Table 3.1. Gross hourly wage gap at the mean and at different quantiles

Mean

	Total		Formal		Informal		
	Mean	sd	Mean	sd	Mean	sd	wf/wi
Overeducated	4627.06	3847.00	5170.34	4116.13	2714.70	1602.93	1.90
Correct	3588.28	2747.15	4125.16	3007.49	2366.05	1409.71	1.74
Undereducated	2665.47	1364.69	3131.68	1443.82	2197.83	1097.70	1.42
Total	**3662.58**	**2894.68**	**4240.56**	**3193.62**	**2379.11**	**1396.24**	**1.78**

Quantiles

Lower - q25

	Total	Formal	Informal	wf/wi
Overeducated	2503.47	2503.47	1944.45	1.29
Correct	2333.33	2503.47	1600.00	1.56
Undereducated	1944.45	2417.59	1555.56	1.55
Total	**2333.33**	**2503.47**	**1633.33**	**1.53**

Middle - q50

	Total	Formal	Informal	wf/wi
Overeducated	3111.11	3402.78	2434.78	1.40
Correct	2700.35	3004.17	2187.50	1.37
Undereducated	2503.47	2654.46	2097.62	1.27
Total	**2722.22**	**3004.17**	**2216.67**	**1.36**

Higher - q75

	Total	Formal	Informal	wf/wi
Overeducated	5185.19	6003.47	2986.67	2.01
Correct	3888.89	4375.00	2722.22	1.61
Undereducated	3004.17	3402.78	2561.36	1.33
Total	**3888.89**	**4612.03**	**2731.06**	**1.69**

Notes: Gross hourly wage in pesos. 'wf ' denotes wages of formal workers and 'wf ' wages for informal workers. 'sd' denotes standard deviation.

Table 3.2. Distribution (%) of workers across years of overeducation, require education and under education by years of actual education

Actual years of education	% of workers	Years of undereducation				Years of overeducation			
		<=-3	-2	-1	0	1	2	>=3	Total
1. Formal workers									
9 or fewer	18.16	30.85	0.29	0.00	68.86	0.00	0.00	0.00	100
10-11	39.26	1.91	0.03	0.00	87.31	0.00	0.03	10.73	100
12-13	12.38	2.47	0.09	0.00	73.74	0.00	0.00	23.70	100
14-15	11.93	0.35	1.06	0.00	70.71	0.00	1.68	26.19	100
16-17	15.27	0.00	0.00	0.00	69.52	0.00	0.21	30.27	100
18 or more	3.37	0.00	0.00	0.00	29.58	8.45	15.49	46.48	100
Total	100	6.66	0.12	0.13	75.85	0.25	0.71	16.28	100
2. Informal workers									
9 or fewer	47.29	29.91	0.15	0.00	69.84	0.00	0.00	0.10	100
10-11	37.44	1.31	0.12	0.00	85.04	0.00	0.00	13.53	100
12-13	6.96	6.04	0.00	0.00	61.41	0.00	0.00	32.55	100
14-15	4.51	0.52	0.52	0.00	55.44	0.00	2.07	41.45	100
16-17	3.36	0.00	0.00	0.00	58.33	0.00	0.00	41.67	100
18 or more	0.44	0.00	0.00	0.00	25.00	0.00	25.00	50.00	100
Total	100	15.07	0.14	0.02	73.74	0.00	0.19	10.84	100

Notes: Figures are in percentages

Table 3.3. Descriptive statistics for the main variables in the analysis

	Total		Formal		Informal	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Gross hourly wage (pesos)	3662.58	2894.68	4240.56	3193.62	2379.11	1396.24
Educational Attainment						
Basic Primary or below	0.14	0.34	0.09	0.28	0.25	0.43
Basic secondary	0.13	0.34	0.09	0.29	0.22	0.42
Secondary	0.37	0.48	0.38	0.48	0.36	0.48
Higher education or more	0.36	0.48	0.44	0.50	0.16	0.37
Education (years)	10.86	3.82	11.73	3.56	8.92	3.65
Age (years)	33.83	10.23	34.64	9.73	32.03	11.03
Experience (years)	17.97	11.47	17.91	11	18.11	12.45
Tenure (months)	47.75	66.21	57.7	72.7	25.67	40.93
Women	0.43	0.49	0.44	0.5	0.41	0.49
Married	0.52	0.5	0.55	0.5	0.46	0.5
Household head	0.43	0.49	0.45	0.50	0.38	0.48
Occupation						
Unskilled	0.31	0.46	0.26	0.44	0.43	0.5
Professionals and Technicians 1	0.07	0.25	0.09	0.28	0.02	0.13
Professionals and Technicians 2	0.04	0.2	0.05	0.22	0.03	0.18
Managers and Public Officials	0.03	0.17	0.03	0.18	0.02	0.13
Administrative Staff	0.21	0.4	0.24	0.43	0.12	0.33
Merchant and Vendor	0.16	0.37	0.15	0.36	0.18	0.39
Service Worker	0.18	0.39	0.18	0.38	0.20	0.4
Type of contract						
No contract	0.29	0.08	0.05	0.06	0.82	0.43
Permanent	0.48	0.50	0.65	0.48	0.10	0.3
Temporal	0.23	0.42	0.30	0.46	0.08	0.27
Hours of work (per week)	50.54	10.59	49.96	9.17	51.82	13.13
Firm size						
Micro (1-10 workers)	0.33	0.47	0.14	0.35	0.74	0.44
Small (11 - 50 workers)	0.2	0.4	0.21	0.41	0.16	0.37
Medium (51- 100 workers)	0.06	0.23	0.08	0.26	0.02	0.14
Large (101 workers or more)	0.42	0.49	0.57	0.49	0.08	0.27
Sector						
Mining, electricity, gas and water	0.03	0.16	0.03	0.18	0.01	0.11
Industry	0.23	0.42	0.23	0.42	0.22	0.42
Construction	0.07	0.26	0.04	0.21	0.13	0.34
Sales, Hotels and Restaurants	0.29	0.45	0.24	0.43	0.41	0.49
Transportation	0.09	0.28	0.1	0.29	0.07	0.25
Financial Intermediation	0.12	0.32	0.15	0.35	0.06	0.23
Social Services	0.18	0.38	0.21	0.41	0.1	0.31
Observations	13797		9513		4284	

Notes: Figures are in percentages, excepting gross hourly wage, education, age, experience and tenure, whose units of measurement are indicated in parenthesis.

Table 3.4. Returns to years of education. Mincer and ORU models

	Mincer				ORU - Mean			
	[1]		[2]		[1]		[2]	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.1008**	0.0543**	0.0900**	0.0419**	-	-	-	-
	[0.0014]	[0.0023]	[0.0014]	[0.0021]				
Surplus	-	-	-	-	0.0931**	0.0416**	0.0860**	0.0362**
					[0.0028]	[0.0052]	[0.0025]	[0.0045]
Required	-	-	-	-	0.1323**	0.0763**	0.1206**	0.0633**
					[0.0017]	[0.0034]	[0.0016]	[0.0035]
Deficit	-	-	-	-	-0.0336**	-0.0468**	-0.0310**	-0.0362**
					[0.0035]	[0.0044]	[0.0032]	[0.0039]
Observations	9512	4284	9512	4284	9512	4284	9512	4284
F-statistic	1014.1	125.5	284.26	72.61	996.3	106.1	328.2	71.2
R squared (adj.)	0.39	0.15	0.50	0.36	0.46	0.16	0.55	0.37

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

Table 3.5. Estimates of the ORU models of earnings with different reference groups for calculating required years of education

	[1]		[2]		[3]	
	Formal	Informal	Formal	Informal	Formal	Informal
Surplus	0.0860**	0.0362**	0.0908**	0.0395**	0.0908**	0.0360**
	[0.0025]	[0.0045]	[0.0031]	[0.0040]	[0.0031]	[0.0057]
Required	0.1206**	0.0633**	0.1281**	0.0699**	0.1281**	0.0582**
	[0.0016]	[0.0035]	[0.0017]	[0.0038]	[0.0017]	[0.0035]
Deficit	-0.0310**	-0.0362**	-0.0355**	-0.0310**	-0.0355**	-0.0383**
	[0.0032]	[0.0039]	[0.0026]	[0.0044]	[0.0026]	[0.0031]
Observations	9512	4284	9512	4284	9512	4284

Notes:

[1] Reference group formal and informal workers together.

[2] Reference group formal and informal workers separately.

[3] Reference group formal workers.

Experience (and its square), tenure (and its square), gender, marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ $p<0.1$, * $p<0.05$, ** $p<0.01$.

Table 3.6. Returns to years of education at the mean and at various quantiles

	OLS		QR					
			25		50		75	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0900**	0.0419**	0.0461**	0.0323**	0.0771**	0.0321**	0.0999**	0.0339**
	[0.0014]	[0.0021]	[0.0009]	[0.0029]	[0.0017]	[0.0021]	[0.0025]	[0.0018]
Surplus	0.0860**	0.0362**	0.0446**	0.0298**	0.0710**	0.0323**	0.0963**	0.0306**
	[0.0025]	[0.0045]	[0.0019]	[0.0058]	[0.0026]	[0.0039]	[0.0034]	[0.0036]
Required	0.1206**	0.0633**	0.0685**	0.0473**	0.1081**	0.0501**	0.1375**	0.0565**
	[0.0016]	[0.0035]	[0.0011]	[0.0044]	[0.0016]	[0.0030]	[0.0023]	[0.0029]
Deficit	-0.0310**	-0.0362**	-0.0232**	-0.0307**	-0.0223**	-0.0261**	-0.0188**	-0.0281**
	[0.0032]	[0.0039]	[0.0025]	[0.0051]	[0.0032]	[0.0033]	[0.0039]	[0.0031]
Observations	9512	4284	9512	4284	9512	4284	9512	4284

Notes: Experience (and its square), tenure (and its square), gender, marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.7. Returns to years of education. Mincer and ORU models - Correcting for selection

	Mincer				ORU - Mean			
	Without		With Selection		Without		With Selection	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0900**	0.0419**	0.0907**	0.0413**	-	-	-	-
	[0.0014]	[0.0021]	[0.0017]	[0.0027]				
Surplus	-	-	-	-	0.0860**	0.0362**	0.0852**	0.0367**
					[0.0025]	[0.0045]	[0.0027]	[0.0048]
Required	-	-	-	-	0.1206**	0.0633**	0.1205**	0.0632**
					[0.0016]	[0.0035]	[0.0017]	[0.0038]
Deficit	-	-	-	-	-0.0310**	-0.0362**	-0.0337**	-0.0359**
					[0.0032]	[0.0039]	[0.0033]	[0.0042]
Mills lambda	-	-	0.2458**	0.0082	-	-	0.1827**	-0.0200
			[0.0462]	[0.0598]			[0.0446]	[0.0572]
Observations	9512	4284	12981	13078	9512	4284	12981	13078

Notes: Experience (and its square), tenure (and its square), gender, marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.8. Returns to years of education at the mean and at various quantiles – Correcting for selection

	OLS		QR					
			25		50		75	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0907**	0.0413**	0.0489**	0.0367**	0.0802**	0.0373**	0.1057**	0.0332**
	[0.0017]	[0.0027]	[0.0008]	[0.0037]	[0.0016]	[0.0020]	[0.0029]	[0.0029]
Mills lambda 1	0.2458**	0.0082	0.6495**	-0.4065*	0.7126**	-0.3671**	0.7524**	-0.1255
	[0.0462]	[0.0598]	[0.0525]	[0.1777]	[0.0892]	[0.0970]	[0.1456]	[0.1423]
Mills lambda 2	-	-	-0.1419**	0.0454	-0.1122**	0.0495*	-0.1109*	0.0503
			[0.0177]	[0.0359]	[0.0310]	[0.0206]	[0.0515]	[0.0309]
Surplus	0.0852**	0.0367**	0.0488**	0.0363**	0.0587**	0.0398**	0.0987**	0.0320**
	[0.0027]	[0.0048]	[0.0019]	[0.0069]	[0.0031]	[0.0039]	[0.0036]	[0.0040]
Required	0.1205**	0.0632**	0.0720**	0.0516**	0.0969**	0.0557**	0.1433**	0.0590**
	[0.0017]	[0.0038]	[0.0011]	[0.0055]	[0.0017]	[0.0031]	[0.0025]	[0.0034]
Deficit	-0.0337**	-0.0359**	-0.0277**	-0.0337**	-0.0049	-0.0312**	-0.0279**	-0.0281**
	[0.0033]	[0.0042]	[0.0024]	[0.0062]	[0.0037]	[0.0034]	[0.0041]	[0.0035]
Mills lambda 1	0.1827**	-0.0200	0.6735**	-0.4087*	0.5327**	-0.4296**	0.8644**	-0.1977+
	[0.0446]	[0.0572]	[0.0600]	[0.1919]	[0.0421]	[0.1089]	[0.1100]	[0.1137]
Mills lambda 2	-	-	-0.1491**	0.0501	-0.2798**	0.0618**	-0.2020**	0.0602*
			[0.0203]	[0.0386]	[0.0253]	[0.0230]	[0.0391]	[0.0247]
Observations	8955	3997	8955	3997	8955	3997	8955	3997

Notes: Experience (and its square), tenure (and its square), gender, marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.9. Implied payoffs to schooling, adjusting for required, over and under education

	Payoff (%)
(A) Adjusted for req-over and under education	
Formal workers	9.06%
Informal workers	4.44%
(a) assuming same earnings effects to require, over and under education	7.45%
(b) as for (a) but also same level of require, over and under education within each schooling category	7.55%
(c) as for (b) but also assuming same distribution across schooling categories	9.06%
(B) Adjusted for require education	
Formal workers	9.06%
Informal workers	4.44%
(a) assuming same earnings effects to require education	6.16%
(b) as for (a) but also same level of require education within each schooling category	6.81%
(c) as for (b) but also assuming same distribution across schooling categories	8.41%
(C) Adjusted for overeducation	
Formal workers	9.06%
Informal workers	4.44%
(a) assuming same earnings effects to overeducation	5.83%
(b) as for (a) but also same level of overeducation within each schooling category	5.24%
(c) as for (b) but also assuming same distribution across schooling categories	5.70%
(D) Adjusted for undereducation	
Formal workers	9.06%
Informal workers	4.44%
(a) assuming same earnings effects to undereducation as formal workers	4.35%
(b) as for (a) but also same level of undereducation within each schooling category	4.39%
(c) as for (b) but also assuming same distribution across schooling categories	5.04%
(E) Adjusted for require and overeducation	
Formal workers	9.06%
Informal workers	4.44%
(a) assuming same earnings effects to require and overeducation	7.54%
(b) as for (a) but also same level of require and overeducation within each schooling category	7.61%
(c) as for (b) but also assuming same distribution across schooling categories	9.06%

Table 3.10. Returns to years of education. Mincer and ORU models - Correcting for selection for men and women

Men

	Mincer				ORU - Mean			
	Without		With Selection		Without		With Selection	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0839**	0.0362**	0.0850**	0.0359**	-	-		
	[0.0018]	[0.0027]	[0.0020]	[0.0031]				
Surplus	-	-	-	-	0.0859**	0.0241**	0.0862**	0.0247**
					[0.0034]	[0.0057]	[0.0035]	[0.0060]
Required	-	-	-	-	0.1197**	0.0515**	0.1196**	0.0506**
					[0.0022]	[0.0048]	[0.0023]	[0.0052]
Deficit	-	-	-	-	-0.0249**	-0.0399**	-0.0276**	-0.0400**
					[0.0039]	[0.0048]	[0.0040]	[0.0051]
Mills lambda	-	-	0.2359**	0.0145	-	-	0.1566**	-0.0001
			[0.0565]	[0.0754]			[0.0548]	[0.0752]
Observations	5367	2542	7460	7526	5367	2542	7460	7526

Women

	Mincer				ORU - Mean			
	Without		With Selection		Without		With Selection	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.1003**	0.0516**	0.1003**	0.0506**	-	-	-	-
	[0.0022]	[0.0035]	[0.0024]	[0.0041]				
Surplus	-	-	-	-	0.0860**	0.0569**	0.0839**	0.0574**
					[0.0039]	[0.0074]	[0.0040]	[0.0079]
Required	-	-	-	-	0.1235**	0.0780**	0.1238**	0.0784**
					[0.0024]	[0.0052]	[0.0026]	[0.0058]
Deficit	-	-	-	-	-0.0490**	-0.0296**	-0.0516**	-0.0293**
					[0.0057]	[0.0066]	[0.0059]	[0.0070]
Mills lambda	-	-	0.2518**	0.0036	-	-	0.2505**	-0.0465
			[0.0756]	[0.0836]			[0.0721]	[0.0826]
Observations	4145	1742	5521	5552	4145	1742	5521	5552

Notes: Experience (and its square), tenure (and its square), marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.11. Returns to years of education at the mean and at various quantiles – Correcting for selection for men and women

Men

	OLS		QR					
			25		50		75	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.0850**	0.0359**	0.0430**	0.0306**	0.0760**	0.0311**	0.1002**	0.0293**
	[0.0020]	[0.0031]	[0.0012]	[0.0046]	[0.0024]	[0.0027]	[0.0037]	[0.0031]
Mills lambda 1	0.2359**	0.0145	0.7010**	-0.3354	0.9781**	-0.2941*	1.0493**	-0.0484
	[0.0565]	[0.0754]	[0.0792]	[0.2193]	[0.1453]	[0.1318]	[0.1954]	[0.1534]
Mills lambda 2	-	-	-0.1492**	0.0439	-0.1850**	0.0372	-0.1887**	0.0635*
			[0.0258]	[0.0444]	[0.0484]	[0.0272]	[0.0662]	[0.0315]
Surplus	0.0862**	0.0247**	0.0483**	0.0196**	0.0777**	0.0278**	0.1039**	0.0198**
	[0.0035]	[0.0060]	[0.0021]	[0.0066]	[0.0037]	[0.0044]	[0.0050]	[0.0061]
Required	0.1196**	0.0506**	0.0719**	0.0366**	0.1154**	0.0439**	0.1461**	0.0535**
	[0.0023]	[0.0052]	[0.0013]	[0.0058]	[0.0025]	[0.0039]	[0.0036]	[0.0060]
Deficit	-0.0276**	-0.0400**	-0.0224**	-0.0365**	-0.0260**	-0.0322**	-0.0290**	-0.0281**
	[0.0040]	[0.0051]	[0.0026]	[0.0058]	[0.0042]	[0.0038]	[0.0052]	[0.0052]
Mills lambda 1	0.1566**	-0.0001	0.8299**	-0.3090+	1.0376**	-0.2953*	1.0037**	-0.0887
	[0.0548]	[0.0752]	[0.0736]	[0.1854]	[0.1211]	[0.1244]	[0.1574]	[0.1683]
Mills lambda 2	-	-	-0.1909**	0.034	-0.2384**	0.0424	-0.2464**	0.0596+
			[0.0240]	[0.0368]	[0.0406]	[0.0259]	[0.0531]	[0.0340]
Observations	7460	7526	5057	2388	5057	2388	5057	2388

Notes: Experience (and its square), tenure (and its square), marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.11 continued

Women

	OLS		QR					
			25		50		75	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Actual	0.1003**	0.0506**	0.0620**	0.0545**	0.0898**	0.0500**	0.1119**	0.0403**
	[0.0024]	[0.0041]	[0.0015]	[0.0052]	[0.0022]	[0.0038]	[0.0041]	[0.0044]
Mills lambda 1	0.2518**	0.0036	0.5398**	-0.3133	0.4388**	-0.4966**	0.3101	-0.3392
	[0.0756]	[0.0836]	[0.0862]	[0.2588]	[0.1202]	[0.1915]	[0.2051]	[0.2279]
Mills lambda 2	-	-	-0.1310**	0.0015	-0.0505	0.0736+	0.0016	0.0807
			[0.0318]	[0.0541]	[0.0469]	[0.0422]	[0.0839]	[0.0528]
Surplus	0.0839**	0.0574**	0.0494**	0.0615**	0.0646**	0.0602**	0.0938**	0.0413**
	[0.0040]	[0.0079]	[0.0030]	[0.0129]	[0.0040]	[0.0067]	[0.0046]	[0.0072]
Required	0.1238**	0.0784**	0.0752**	0.0719**	0.1089**	0.0664**	0.1410**	0.0671**
	[0.0026]	[0.0058]	[0.0018]	[0.0092]	[0.0026]	[0.0050]	[0.0031]	[0.0054]
Deficit	-0.0516**	-0.0293**	-0.0445**	-0.0322**	-0.0492**	-0.0320**	-0.0279**	-0.0285**
	[0.0059]	[0.0070]	[0.0045]	[0.0118]	[0.0059]	[0.0062]	[0.0063]	[0.0065]
Mills lambda 1	0.2505**	-0.0465	0.3667**	-0.2733	0.4504**	-0.4044*	0.7210**	-0.3188
	[0.0721]	[0.0826]	[0.0943]	[0.3492]	[0.1251]	[0.1901]	[0.1387]	[0.2079]
Mills lambda 2	-	-	-0.0628+	-0.0055	-0.0847+	0.0479	-0.1213*	0.0666
			[0.0347]	[0.0731]	[0.0481]	[0.0418]	[0.0564]	[0.0484]
Observations	3898	1609	3898	1609	3898	1609	3898	1609

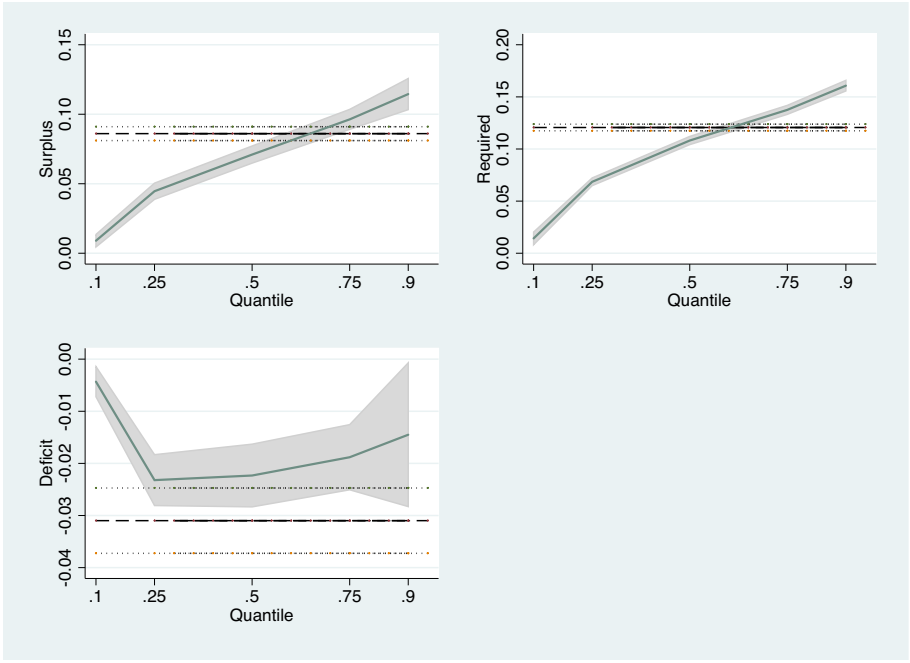
Notes: Experience (and its square), tenure (and its square), marital status, head of household, hours worked, type of contract, size of the firm, sector and region are included as controls in all regressions. Standard errors in [].+ p<0.1, * p<0.05, ** p<0.01.

Table 3.12. Implied payoffs to schooling, adjusting for required, over and under education

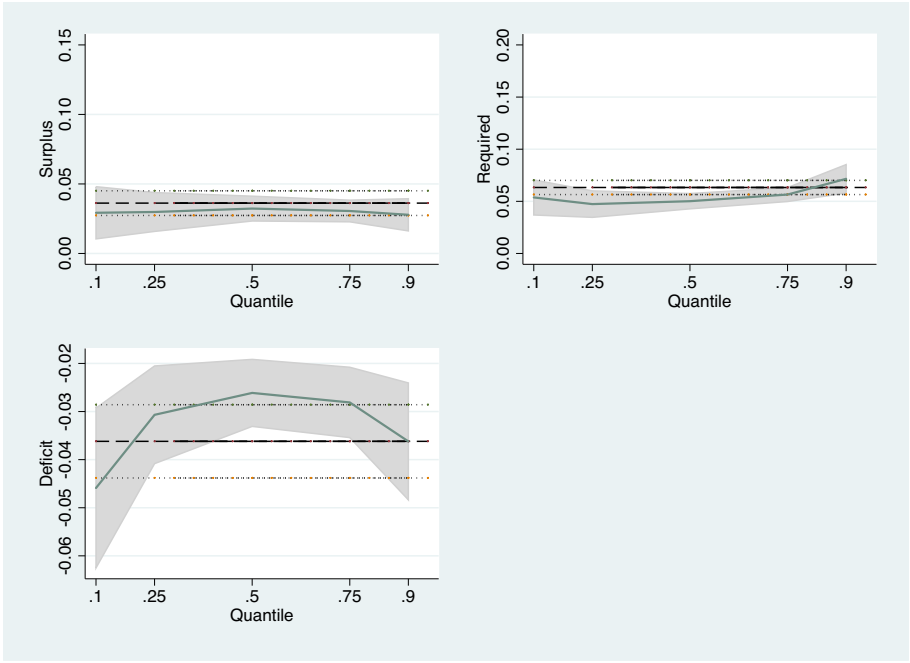
	Payoff (%)	
	Men	Women
(A) Adjusted for req-over and under education		
Formal workers	8.38%	10.21%
Informal workers	3.78%	5.44%
(a) assuming same earnings effects to require, over and under education	6.78%	8.61%
(b) as for (a) but also same level of require, over and under education within each schooling category	6.70%	9.20%
(c) as for (b) but also assuming same distribution across schooling categories	8.38%	10.21%
(B) Adjusted for require education		
Formal workers	8.38%	10.21%
Informal workers	3.78%	5.44%
(a) assuming same earnings effects to require education	5.42%	7.05%
(b) as for (a) but also same level of require education within each schooling category	5.91%	8.52%
(c) as for (b) but also assuming same distribution across schooling categories	7.41%	9.97%
(C) Adjusted for overeducation		
Formal workers	8.38%	10.21%
Informal workers	3.78%	5.44%
(a) assuming same earnings effects to overeducation	5.70%	6.15%
(b) as for (a) but also same level of overeducation within each schooling category	5.03%	5.54%
(c) as for (b) but also assuming same distribution across schooling categories	5.23%	6.41%
(D) Adjusted for undereducation		
Formal workers	8.38%	10.21%
Informal workers	3.78%	5.44%
(a) assuming same earnings effects to undereducation as formal workers	3.22%	6.30%
(b) as for (a) but also same level of undereducation within each schooling category	3.32%	6.02%
(c) as for (b) but also assuming same distribution across schooling categories	3.75%	7.00%
(E) Adjusted for require and overeducation		
Formal workers	8.38%	10.21%
Informal workers	3.78%	5.44%
(a) assuming same earnings effects to require and overeducation	7.34%	7.75%
(b) as for (a) but also same level of require and overeducation within each schooling category	7.16%	8.62%
(c) as for (b) but also assuming same distribution across schooling categories	8.63%	9.79%

Figure 3.1. Returns to surplus-required-deficit years of education over the entire distribution

Formal workers



Informal workers



Chapter 4. Wage Gaps Across Colombian Regions: The Role of Education and Informality

4.1 Introduction

Over the past decade, several studies have registered the decline in income inequality for Latin America countries (López-Calva & Lustig, 2009 and Gasparini, Cruces and Tornarolli, 2011). While this new trend in income inequality has received special attention at the national level, few studies have looked at income inequality at the regional level for Latin America countries. Regional studies are of great relevance, because even in the presence of declining income inequality at national level, important inter-regional disparities may persist. This is so, because socio economic indicators at the national level can often hide significant variances between territories of the same country. This study considers the case of Colombia, a country that despite a decrease in income inequality in the past decade presents one of the highest Gini coefficients of Latin America countries and faces large geographical differences. Colombia shows important disparities in economic and social development among its regions. This implies that an important part of inequality between Colombian individuals may be the consequence of inequality between regions of the country (Bonet and Meisel, 2008; Jourmard and Londoño, 2013). In particular, differences in wages deserve attention from a regional perspective as, for example, in 2010 the average gross hourly wage of a small city, such as Cucuta, is only 66% of that paid in Bogotá.

To explain large spatial wage disparities, several explanations have been proposed. One of them emphasizes that wage differences across areas are caused by differences in amenities. For instance, certain areas may have a favorable climate and more access to natural resources. Under this context, wage differentials may be seen as compensated differentials, meaning that some areas may have higher wages to attract workers so as to compensate for

the lack of amenities (Greenwood et al. 1991). Another explanation is related to the point that differences in wages across regions could reflect spatial differences in the skill composition of the workforce (Combes, Duranton and Gobillon, 2008). Workers with better labor market characteristics tend to sort themselves in areas that concentrate industries with high skill requirements where wages tend to be higher. Associated to this last explanation, the third one is based on agglomeration economies. A larger pool of high skill workers in an area may provide a source of important knowledge spillovers that can lead to productivity gains (Glaeser et al., 1992). Also, labor pooling improves the matching between firms and workers, which could also increase economic efficiency and lead to higher wages (Andersson, Burgess and Lane, 2007).

A number of studies have been devoted at measuring the degree of regional wage gaps and identifying their origin. For instance, Blackaby and Murphy (1995) and Duranton and Monastiriotis (2002) analyze the case of Britain, García and Molina (2002), Motellón, López-Bazo and El-Attar (2011), López-Bazo and Motellón (2012) that of Spain and Pereira and Galego (2013) the one of Portugal. These studies center their analysis on the estimation of human capital wage equations and on decomposition analysis. The decomposition analysis is based on the idea that regional wage differentials are the result of how characteristics that determines wages are distributed across regions (the endowment component) and by how different these characteristics are rewarded across space (the coefficients or wage structure component). The extent to which these two components explain regional wage differentials has been of great interest in past studies and their importance in explaining regional wage gaps differ considerably across and within countries. Some studies conclude that the regional wage differentials are mostly due to differences in individual characteristics between regions (Blackaby and Murphy, 1995). Other studies found that a significant part of wage differentials are explained by difference in returns, (Motellón, López-

Bazo and El-Attar, 2011 and Pereira and Galego, 2013). While some studies point that both components play an important role (García and Molina, 2002).

To the best of our knowledge almost all studies that analyze regional differentials for Colombia and other developing countries are aggregate approaches. These approaches are centered on a single aggregate variable, usually per capita income, at the regional level. For example, Bonet and Meisel (2007) study the convergence in regional income in Colombia, for the administrative division of departments and analyzed the period between the years 1975-2000, concluding that there is a process of polarization. Galvis (2004) using series of average hourly wage for the four major Colombian metropolitan areas and for different educational levels, analyzed if there is market integration, i.e. there is wage convergence. He concluded that there is some convergence in wages for those segments of the labor force that reached primary and secondary education level. However at the highest educational level, there is not convergence for all the cities, probably due to the great heterogeneity of workforce at this segment. Unfortunately, aggregate approaches hardly say anything about what factors explain regional inequalities.

Azzoni and Servo (2002) using micro-data for the 10 largest metropolitan regions in Brazil found that wage differentials were lower after adding controls for worker and job characteristics and cost of living, though they remain sizeable. With regard to the factors that explain regional wage disparities, they found education as the most important variable for explaining such differences. Romero (2008) pursued a similar study for the Colombian case, and concluded that a significant part of regional labor income differences disappeared after adding controls for workers and firms characteristics and cost of living. His results indicate that, the contribution attributed to regional differences in the cost of living is negligible for explaining labor income inequality across regions, while difference in education is the most important

source of the observed regional labor income disparities. Quiñones and Rodriguez (2011) reach the same conclusion after implementing the Blinder (1973) and Oaxaca (1973) decomposition for evaluating the contribution of differences in education in explaining wage differentials across Colombian regions. So it can be concluded from past studies that differences in the endowment of human capital and in its returns has been the most important factor for explaining regional wage differentials.

In this study, as in past studies, special attention is paid to spatial imbalances in the endowment of human capital, and to what extend these differences and the regional heterogeneity in the return to this type of capital may help to explain regional wage gaps. But unlike most previous studies done for developing countries, regional wage gaps are estimated for several quantiles of the wage distribution in order to analyze the contribution of education at different points on the wage distribution. Moreover as a novel and main contribution, this paper will not only focus on the regional differences in the endowments of human capital, but will go further in exploring one important feature of almost all developing countries: the stylized fact that a large proportion of the employed population in Colombia has an informal job. More importantly, recent studies for Colombia have emphasized that informal jobs are not equally distributed across the main metropolitan areas of the country (Galvis, 2012). In Colombia some cities have informality rates of around 60% while others have rates of about 20%. In addition, we build on the results in the study by Ortiz, Uribe and Badillo (2008), which indicates that the Colombian labor market is segmented in two dimensions. An intra-regional or scale segmentation, which is mainly due to the restrictions on the access to physical and human capital that limited the possibility of expansion of firms to a larger scale. This type of segmentation may imply that workers and employers in the informal sector, usually associated with small

establishments[1], face significant barriers in the transition to the formal sector, with higher productivity and higher income. The second type of segmentation is the inter-regional segmentation, which is mainly due to the barriers of mobility of labor and other factors between regions. Accordingly, the hypothesis of our study is that regional wage inequality may be explained by regional differences in the availability of good jobs that generate higher wages. Meaning that, apart from the differences in the endowment of human capital across Colombian regions, regional heterogeneity in the incidence of informality may be another important source of regional wage disparities.

The empirical analysis consists of examining the returns to education and the pay penalty of informal jobs across Colombian regions by using mean models and quantile regression models in order to analyze the effect of characteristics along the wage distribution. Then, regional wage gaps are decomposed into the contribution of differences in the regional distribution of characteristics, and into the contribution of differences in wage structures (heterogeneity in prices to characteristics). In doing so, we apply the standard Blinder-Oaxaca decomposition at the mean and the decomposition for unconditional quantile regression (UQR) models proposed by Firpo, Fortin and Lemieux (2009, 2011) at selected quantiles. With both of these approaches it is possible to isolate the particular contribution of education and informality to the regional wage gap, in contrast with other procedures (Machado and Mata, 2005; Melly, 2005). Pereira and Galego (2013) applied this method in the case of regional wage differentials for Portugal. As far as we know, our study represents the first application of this method for the analysis of regional wage differentials of a developing country.

Results for Colombia show that regions not only differed in earning relevant characteristics, but also display sizeable regional variability in the returns to these characteristics. Particularly, heterogeneity in returns to

[1] However, establishment size and sector assignments have been found to be imperfectly correlated.

education across regions play an important role in explaining regional wage gaps. Additionally, workers face different informal pay penalties throughout the territory and it affects mostly individuals at the lower part of the wage distribution, therefore its contribution in explaining regional wage gaps is limited to this part. Our results confirm previous evidence on the existence of significant regional wage differences between the Golden Triangle region, conformed by the cities of Cali, Medellin and Bogota, and other regions in the country. The difference is particularly wide for those regions with a large share of labor in the informal sector. In fact, after comparing formal workers across regions and separately doing the same for informal workers, regional wage gaps are reduced considerably. Furthermore, our results reveal that not distinguishing between formal and informal workers leads to conclusions on the origin of regional wage disparities that are partially misleading. For instance, the belief that the Golden Triangle is the region with the best endowed workforce is not completely accurate when the analysis distinguishes between formal and informal workers. Moreover, it seems that the distribution of education is generating an equalizing effect of wages across some regions, whereas the returns to education continue to be a source of wage inequality across Colombian territories.

The results of this study point to the conclusion that some public policies aim in reducing human capital differences among regions will help to decrease regional wage gaps, especially at the higher parts of the wage distribution. However, equalizing years of education of workers across regions would not be enough to reduce regional wage differences due to the sizeable differences in returns to years of education at higher quantiles. Similar results have been found in previews studies, albeit in a context of developed countries. Meanwhile policies that points towards the reduction of informality will help to minor regional wage gaps at the lower part of the wage distribution particularly for those regions with sizable informality.

The remainder of the chapter is organized as follows. The next section presents a description of the data used. Section 4.3 outlines the methodology used in this study. Then, sections 4.4 and 4.5 report and discusses the results. Finally, in section 4.6 conclusions are presented.

4.2 Data and descriptive analysis

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

The analysis was restricted to salary workers that were not carrying formal studies aged between 15 and 60 years and who report working more than 16 hours per week. We do not include self-employed and employers workers in the analysis because their source of income is a combination of labor and physical capital and therefore may not be compared with earnings of other employees. Apart from this, self-employed workers' earnings would be expected to have a greater measurement error. We also exclude public employees from the sample since public wages are fixed at the national level for all the public administration along the territory so that the regional wage differentials may be artificially lower if public employees are included in the analysis. After excluding observations with missing values or inconsistencies for the selected regressors, 13796 individuals remained in our sample.

Given the importance of labor market inequality dynamics in explaining the trend in inequality, and since earnings obtained in the labor market are the main sources of income; this chapter will be focus on analyzing wage

inequality at the regional level. We have combined information from gross monthly income and worked hours in order to obtain gross hourly wages. A first look at the degree of regional wage differentials in Colombia is obtained from a simple inspection of Table 1, which in the second column displays the average gross hourly wage. Large differences in average wages across the thirteen metropolitan areas are observed. For instance, the average wage in Cucuta, the metropolitan area with the lowest level, was 66.15% of the average wage in Bogotá, the metropolitan area with the highest level. As in previous studies, we attempt to control for price differentials by adjusting the nominal gross hourly wage using the deflator from the consumer price index of each city. Consumer price indices for the main city of each metropolitan area were obtained from DANE. We applied the consumer prices index of the main city to the whole metropolitan area. This implies that the price level of the main city is representative for the whole metropolitan area. The averages of this *adjusted* gross hourly wages are shown in the third column of Table 4.1. It is observed that the position in the regional ranking of wages is fairly the same and that the metropolitan areas in the top and the bottom of the ranking remain unchanged. The fact that the consumer price index is built with a base year fairly recent, 2008, may explain the small variation obtained after controlling for difference in prices across the metropolitan areas. However, as far as we know this is the only information on regional prices available for Colombia.

The regional wage gap observed may be caused because worker´s characteristics differ across the metropolitan areas. In particular, they are known to differ in the workers' endowment of education, which is one of the essential determinants of wages. Table 4.1 contains the average years of education of workers for each metropolitan area. As it can be seen, there are notable differences in education. On average, workers in Cartagena have more than two years of education than those workers in Cucuta. On the other hand,

as has already been mentioned, past studies for Colombia have show that the incidence of informality across regions is remarkably different. Since informal workers earn considerably lower wages than their formal counterparts, then a metropolitan area with a higher proportion of informal workers may have lower wages than a metropolitan area with a low fraction of informal workers. As in previous chapters, we define workers as formal if they contribute both to health and old-age insurance. Table 4.1 also presents the percentage of informal workers in each of the metropolitan areas. In accordance with what has been found in previous studies, the incidence of informality is very different across the metropolitan areas. While Cucuta displays an informality of around 59%, the share of informal workers in Medellin is about 19%. Interestingly, some metropolitan areas with the lower average hourly wages are also those with the highest levels of informality (Villavicencio, Pasto and Cucuta). So these simple descriptive figures suggest a negative correlation between the incidence of informality and the hourly wages in the Colombian metropolitan areas.

In order to make the analysis more tractable and for seek of brevity, metropolitan areas where grouped into regions. In Colombia, six regions have been delimited by geographical proximity and natural characteristics (climate, mountains, proximity to the sea, etc…). According to DANE Colombia is delimited into nine regions: Atlantic, Oriental, Central, Pacific, Bogota, Antioquia, Valle del Cauca, San Andres and Providencia and Orinoquia – Amazonia. Though Bogotá, Antioquia and Valle del Cauca belong to one of the six regions, according to the geographical and natural delimitation, they are taken away from their corresponding region because of their economic importance. In our particular case, we grouped the largest metropolitan areas of these regions (Bogotá, Medellin and Cali, correspondingly) into one region

that we will refer to as the Golden Triangle[2]. These metropolitan areas are the most dynamic and productive of the country. The most productive firms, most of the R&D investment executed in the country and the highest skill workers are concentrated in these three areas. Although CHS (2010) does not contain information about the areas of San Andres and Providencia and Orinoquia-Amazonia, there is at least one metropolitan area for each of the remaining regions. Therefore, according to geographical, natural and economic factors we have grouped the metropolitan areas in the dataset into five regions. The first region, Atlantic, includes Barranquilla, Cartagena and Monteria. The second region, Oriental, groups Cucuta, Bucaramanga, and Villavicencio. The third one, Central, is represented by Manizales, Pereira and Ibague, and the fourth, Pacific, is only composed by Pasto. Finally, the fifth region, Golden Triangle, is composed by the three largest metropolitan areas of Colombia, Bogotá, Medellin and Cali.

Table 4.2 provides a description of hourly wages for the five regions. Clearly, average hourly wages differ between regions, although the magnitude of the differences is lower than the one found for the thirteen metropolitan areas. Now, the average hourly wage of the region with the lowest level, Pacific, is 74% of that in the region with the highest level, Golden Triangle. So by grouping metropolitan areas into regions the amount of disparities is attenuated, but they still remain sizable. Apart from the differences in the mean, the wage distributions of these five regions present other interesting variations. For instance, Table 4.2 shows that the wage distributions of the regions have different degree of dispersion. The standard deviation of the logarithm of gross hourly wages and the Gini index for the region with the lowest level of wages, Pacific, are higher than that of the region with high level

[2] Colombia's Golden Triangle refers to an urban region, limited by a triangle whose vertexes are defined by the three largest cities: Bogotá, Medellin and Cali. In our particular case, we are not referring to the region, but only to the three cities that demarcates the triangle.

of wages, Golden Triangle, suggesting that regions also differ in terms of the amount of intra-regional inequality. Finally, from the value of hourly wages at certain percentiles (25%, 50% and 75%)[3], reported in the last columns of Table 4.2, it can be concluded that regional wage differentials are far from constant over the entire wage distribution, with symptoms of a non-monotonic behavior.

In order to have a better comparison of the entire wage distributions Figure 4.1 displays kernel density estimates for hourly wage distributions of the thirteen metropolitan areas and divided into the five regions. Though in particular cases the distribution of hourly wages behaves quite different across the metropolitan areas that comprise each region, in general terms the differences within each region are limited. In fact, it was expected that some heterogeneity in term of wages and other characteristics remains for some regions, as the grouping criteria not always obeyed to economics factors. On the other hand, Figure 4.2 displays kernel density estimates for hourly wage distributions once the metropolitan areas are grouped into regions. As it can be seen there are differences in the shape of these distribution. Noticeably, Pacific stands as the region with the higher wage dispersion; its density lies to the left of other regions and displays a highest mass of probability in the lower tail (larger percentage of workers with lower wages). Oriental and Central regions have a similar pattern as Pacific but less discernible. Hourly wage densities of the Atlantic region and the Golden Triangle are slightly to the right of the rest of other regions and have a narrower left tail. So, the evidence from Table 4.2 and Figure 4.3 confirms that there are noticeable differences across regions in the entire wage distribution, and not just on average wages.

[3] In order to save space we do not reproduce here the results in this chapter for other percentiles, although they are available upon request. In any case, including results corresponding to more percentiles does not modify the general conclusions regarding regional disparities over the entire wage distribution.

To account for these differences, in the rest of this Chapter we provide results for the average and for some selected quantiles.

As it has been already mentioned, some part of the regional wage differentials might be caused by the spatial distribution of human capital and other earning relevant determinants, as informality. To explore this event, Table 4.3 reports a simple description of the observable worker and firm characteristics for the five regions. It is for instance observed that regions with high levels of wages have workers employed in relatively larger firms and with a permanent contract. Other differences are worth examining more closely. For example, the proportion of workers employed in the sectors of industry and financial intermediation is larger in high wage regions. One point that also worth to mentioning is the low proportion of women working in Atlantic region, 39%, compare to 45% in Golden Triangle. Informality also differs considerably between regions; the incidence of informality is 49% in Pacific while in the Golden Triangle is 23%. These differences in the proportion of informal workers across regions might intensify regional wage differentials, since formal jobs usually entail higher wages than informal jobs. Hence the wage distribution of Pacific region might be concentrated in the lower tail because this region displays the highest incidence of informality.

Therefore, there are differences in characteristics between regions that may result in regional wage differentials. Nevertheless the key point is if differences in characteristics can mainly account for regional wage differences, or if part of the wage gap is produced by differences in how these characteristics are paid across regions. If regional wage gaps were completely explained by differences in the distribution of observable characteristics across regions, then under such circumstances, similar workers employed in similar firms but located in different regions would earn the same wage. On the contrary, if part of the wage gap could be explained by differences in how characteristics are rewarded, this could be associated to failures in regional

labor markets, as similar workers in comparable firms but in different regions would be earning different wages. In the section that follows we aim to shed more light on this issue.

4.3 Empirical strategy

4.4.1 Specification of the wage equation

The empirical strategy is based on a model in which the wage of individual *i* in region *r* is given by:

$$W_{ir} = X_{ir}\boldsymbol{\beta}_r + \varepsilon_{ir} \quad (4.1)$$

where W_{ir} denotes the log of the hourly wage of individual *i* in region *r*. X_{ir} denotes the set of characteristics that affect the wage of this individual, including years of education, experience (and its square), tenure (and its square), gender, sector of employment, marital status, head of household, hours worked, type of contract, size of the firm and firm sector. $\boldsymbol{\beta}_r$ is the vector of prices or returns at region *r* associated to the characteristics in X_{ir}.. Equation (4.1) is estimated for each region, so that an estimate of the effect of education and informality is obtained for each region rather than imposing the same effect for all regions. This is different from what was done in previous studies for Colombia (Romero, 2006; Quiñones and Rodriguez, 2011), where a dummy variable for each region is introduced thus imposing the same return for workers' and firms' characteristics for all regions. This is a restrictive assumption; if there is inter-regional segmentation then workers with similar characteristics may obtain different returns across regions, not only for education but also for other relevant characteristics that determine wages.

There are two potential sources of bias when estimating equation (4.1). One is related with the sample selection on wages caused by the probability of employment. It arises because some unobserved characteristics could be correlated with the likelihood of employment and wages. Another source of sample selection comes from the probability of being a migrant, as for example, there could be unobservable spatially factors that affect both the probability of migrate and that can be correlated with wages. Although both sources of selection may lead to biased results, there are two reasons why they are not addressed in this chapter. The first one is that previous studies that have attempted to solve for the employment selection have found that the results are not strongly affected by this type of selection (Quiñones and Rodriguez, 2011). On the other hand, internal migration in Colombia has been found to be relatively low, so that this source of selection does not seem to be especially relevant (Ortiz, Uribe and Badillo, 2008).

The analysis from equation (4.1) is based on the mean. However, the descriptive in the previous section showed that regional disparities are far from uniform over the entire wage distribution. Therefore, it is of interest to know the effects of the exogenous variables, for example education, at different points of the distribution of wages. This can be done by using the conditional quantile regression (CQR) model introduced by Koenker and Bassett (1978). It can be written as:

$$W_{ir} = X_{ir}\boldsymbol{\beta}_{\tau r} + \varepsilon_{\tau ir} \text{ with } Q_{\tau}(W_{ir}|X_{ir}) = X_{ir}\boldsymbol{\beta}_{\tau r} \quad (4.2)$$

where $Q_{\tau}(W_{ir}|X_{ir})$ denotes the τ-th conditional quantile of wages given the set of characteristics in X_{ir}. Analogous to the OLS regression of W_{ir} on X_{ir}, where $\boldsymbol{\beta}_r$ is estimated as a solution of minimizing sum of square residuals, $\boldsymbol{\beta}_{\tau r}$ associated with τ-th conditional quantile function may be estimated by

minimizing a sum of asymmetrically weighted absolute residuals (Koenker, 2005; Koenker and Bassett, 1978):

$$\min_{\beta_{\tau r}} \sum \rho_{\tau r}(W_{ir} - X_{ir}\boldsymbol{\beta}_{\tau r}) \tag{4.3}$$

where $\rho_{\tau r}$ is the absolute value function defined as:

$$\rho_{\tau}(u) = u \cdot (\tau - I(u < 0) \text{ for any value } \tau \in (0,1) \tag{4.4}$$

The estimated coefficients of $\boldsymbol{\beta}_{\tau r}$ may be interpreted as marginal or partial effects (depending on whether the corresponding covariate is continuous or binary) on the conditional quantile of interest. If $\boldsymbol{\beta}_{\tau r}$ is a consistent estimator of the conditional and unconditional quantile of W_r, then the underlying data generating process follows a linear-in-parameters additive model structure, i.e. is a pure parallel location-shift data generating process for every covariate. However if the conditional effect of a specific exogenous variable in X_r varies over levels of other exogenous variables in X_r, $\boldsymbol{\beta}_{\tau r}$ may be a consistent estimator of the conditional effect of an exogenous variable at the mean values of the other *k-1* remaining exogenous variables, but is not a consistent estimator of the unconditional effect of X_r (Borah and Basu, 2013). Meaning that, for example, the 90th percentile of the unconditional distribution of wages may not be the same as the 90th percentile of wages conditional on years of education.

It is possible to estimate the unconditional quantile effect of X_r using the approach proposed by Firpo, Fortin and Lemieux (2009) based on the influence function (IF) and recentered influence function (RIF). In the context of wages, the IF is:

$$IF(W_r; q_{\tau}) = (\tau - I\{Y \leq q_{\tau}\})/f_{W_r}(q_{\tau}) \tag{4.5}$$

where q_τ refers to the τ-th unconditional quantile of wages, $f_{W_r}(q_\tau)$ is the probability density function of W_r evaluated at q_τ, and $I\{Y \leq q_\tau\}$ is an indicator variable to denote whether an outcome value is less than q_τ or not. By definition the RIF is equal to:

$$RIF(W_r; q_\tau) = q_\tau + IF(W_r; q_\tau) \tag{4.6}$$

Firpo, Fortin and Lemieux (2009), demonstrate that the implementation of the UQR is straightforward and similar to the OLS regression. For a specific quantile τ, the first step is to estimate the RIF of the τ-th quantile of W_i following eq. (4.5) and eq. (4.6). The second step is to run OLS regression of the $RIF(W_{ir}; q_\tau)$ on the observed covariates, X_{ir}.

$$E[RIF(W_{ir}; q_\tau|\mathrm{X}_{ir})] = X_{ir}\boldsymbol{\beta_{\tau r}} \tag{4.7}$$

Coefficients $\boldsymbol{\beta}_{\tau r}$ represents the approximate marginal effects of the explanatory variables on the unconditional quantile q_τ of wages for workers in region *r*.

4.4.2 Decomposition of regional wage gaps

The Blinder-Oaxaca type decomposition is formulated for decomposing mean differences in log wages between two groups after the estimation of the wage equation (4.1).[4] In our particular case, the wage gap between a high wage region (*r=h*) and a low wage region (*r=l*) can be specified as:

[4] The Blinder-Oaxaca decomposition has been applied in several studies analyzing gender, black and white, public and private wage gaps, and recently it has also been applied for understanding regional

$$\overline{W_h} - \overline{W_l} = (\bar{X}_h - \bar{X}_l)\widehat{\boldsymbol{\beta}}_h - \bar{X}_l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{h}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{l}}) \quad (4.8)$$

The first term corresponds to the differences in the average values of observed workers and firms' characteristics between regions *h* and region *l*, and the second term, is the part of the wage gap attributable to differences in the estimated coefficients or differences in the wage structure.

It is possible to obtain a decomposition of the wage differential at quantile τ, similar to the classical Blinder-Oaxaca decomposition, for any two regions using the RIF regression approach by Firpo, Fortin and Lemieux (2009). Any distributional parameter, for example a wage quantile, can be written as a function $q_\tau(F_W)$ of the cumulative distribution of wages, $F_W(W)$. For example the difference in a wage quantile τ, Δ^{q_τ}, between a high wage region ($r=h$) and a low wage region ($r=l$), can be written as:

$$\Delta^{q_\tau} = q_\tau\left(F_{w_h|r=h}\right) - q_\tau\left(F_{w_l|r=l}\right) \quad (4.9)$$

$$\Delta^{q_\tau} = \left[q_\tau\left(F_{w_h|r=h}\right) - q_\tau\left(F_{w_l|r=h}\right)\right] + \left[q_\tau\left(F_{w_l|r=h}\right) - q_\tau\left(F_{w_l|r=l}\right)\right]$$

$$\Delta^{q_\tau} = \Delta_S^{q_\tau} + \Delta_X^{q_\tau}$$

where $q_\tau\left(F_{w_h|r=h}\right)$ indicates the actual wage quantile of workers belonging and rewarded under the wage structure of region $r=h$. $q_\tau\left(F_{w_l|r=h}\right)$ represents the counterfactual wage quantile, the wage quantile that would prevailed if workers observed in the region with high wages, $r=h$, had been paid under the wage structure of workers in the low wage region, $r=l$. Using the actual and counterfactual wage quantile for each region it is possible to decompose the

wage gaps (García and Molina, 2002; Motellón, López-Bazo and El-Attar, 2011; Pereira and Galego, 2013).

wage gap at any quantile, Δ^{q_τ}, in two terms, one which captures the wage structure effect, $\Delta_S^{q_\tau}$, and another that represents the endowments effect $\Delta_X^{q_\tau}$.

However, as in the case of the Blinder-Oaxaca decomposition for the mean, if the true conditional expectation is not linear, the decomposition based on a linear regression may be biased (Barsky et al., 2002). A reweighted procedure and the RIF-regressions can solve this problem (Firpo, Fortin and Lemieux, 2007, 2011). First a reweighting factor has to be calculated in the following way:

$$\Psi(X) = \frac{\Pr(r = h|\mathrm{X}) / \Pr(r = h)}{\Pr(r = l|\mathrm{X}) / \Pr(r = l)} \tag{4.10}$$

Then RIF-regressions are computed for workers in regions *l*, *h* and for the counterfactual l^c region, using the weights in $\Psi(X)$, to later calculate the next decomposition:

$$\hat{\Delta}^{q_\tau} = \left(\bar{X}_h \widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau h}} - \bar{X}_l^c \widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}}\right) + \left(\bar{X}_l^c \widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}} - \bar{X}_l \widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}\right) \tag{4.11}$$

$$\hat{\Delta}^{q_\tau} = \quad \hat{\Delta}_S^{q_\tau} \quad + \quad \hat{\Delta}_X^{q_\tau}$$

where $\bar{X}_r, r = l, h$, denote the mean wages in regions *l* and *h*, and $\bar{X}_l^c$ is the counterfactual mean for region *l* using the reweighting factor in (4.9) so to make the distribution of the characteristics, *X*, in the region with low wages similar to that of region with high wages.

The wage structure effect can be divided into a pure wage structure effect and a component measuring the reweighting error, as follows:

$$\hat{\Delta}_S^{q_\tau} = \bar{X}_h(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau h}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}}) + (\bar{X}_h - \bar{X}_l^c)\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}} \tag{4.12}$$

$$\hat{\Delta}_S^{q_\tau} = \quad \hat{\Delta}_{S,p}^{q_\tau} \quad + \quad \hat{\Delta}_{S,e}^{q_\tau}$$

The reweighting error goes to zero as $\bar{X}_l^c \longrightarrow \bar{X}_h$.

Similarly, the composition effect can be divided into a pure composition effect and a component for the specification error as:

$$\hat{\Delta}_X^{q_\tau} = (\bar{X}_l^c - \bar{X}_l)\hat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}} + \bar{X}_l \ (\hat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}^{\boldsymbol{c}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{\tau l}}) \tag{4.13}$$

$$\hat{\Delta}_X^{q_\tau} = \quad \hat{\Delta}_{X,p}^{q_\tau} \quad + \quad \hat{\Delta}_{X,e}^{q_\tau}$$

4.4 Results

4.4.1 OLS estimates of the wage equation

Table 4.4 reports the results of estimating Mincer wage equations by OLS and by quantile regressions (conditional and unconditional) for different quantiles for the five regions. Since the particular focus of this chapter is on the effect of education and informal work, the results are shown only for the estimates of the coefficients associated to years of education and informality, though a large set of controls was included as regressors.[5] The first column in Table 4.4 contains the estimates in the mean, that is to say the results of the OLS estimates. The estimated returns to schooling for each region are displayed in the upper panel of the table. As expected, there are significant differences in returns to years of education between regions. For example, a higher return to schooling is observed in those regions with the highest levels of wages. The returns to schooling in Atlantic and Golden Triangle are 8.14% and 8.26% respectively. On the other hand, those regions with the lowest levels of hourly wages display the lowest returns to schooling, 5.57% in the Oriental region and 6.82% in Pacific. Thus, in addition to differences in the

[5] An appendix with the full set of estimates is available upon request.

endowment of education, returns to schooling may be thought to be an important factor in explaining wage gaps across regions.

The OLS estimates of the informal pay penalty, reported in the lower panel of Table 4.4, show a more complex pattern. The Pacific region is the one with the higher pay penalty; informal workers earn 26.8% less than their formal counterparts. However the next region with the higher pay penalty is the Golden Triangle, with a 13.56%. Even though the pay penalty is considerably larger in the region with the lower level of wages compare to the region with the highest level of wages, there seems not to be a clear pattern between informal pay penalty and regional wage gap, when comparing for example the Golden Triangle with the Oriental region, since the pay penalty in the last region is lower. Therefore, the OLS results indicate that Colombian region differ not only in the incidence of informality (the share of the informal sector) but also in the difference in mean wages earned by otherwise similar formal and informal workers.

4.4.2 Quantile regression estimates of the wage equation

Table 4.4 also displays the results of estimating the Mincer wage equations by conditional and unconditional quantile regressions. Results concerning conditional quantiles show conditional returns to schooling and pay penalty earnings after adjusting for workers' and firms' characteristics. Information about the dispersion of wages within groups of individuals with the same characteristics can be derived from the CQR results. Consistently with previous literature, returns to schooling are heterogeneous and increasing along the quantiles for all regions. CQR results suggest that in the Golden Triangle region returns to schooling range from 4.62% for the first quantile to 8.99% for the last quantile of the conditional distribution of wages. While in the Pacific region, returns to schooling in the first quantile are 5.16% and in

the last quantile are 7.29%. Interestingly, the returns to schooling are higher for Pacific compare to those in Golden Triangle at lower quantiles, they are fairly the same at the middle part of the distribution, and lower at higher quantiles. The coefficient of years of education increases along the wage distribution for all regions, suggesting that increasing education has an unequalizing effect in the conditional wage distribution. This unequalizing effect is especially strong for Golden and Atlantic regions, where returns to education increase substantially across the lowest and the highest quantile.

However interpreting conditional quantile regression results must be done cautiously. A common difficulty associated with interpreting these results is that, as has already been mentioned, the 90th percentile of the unconditional distribution of wages may not be the same as the 90th percentile of the conditional distribution of wages. Then the positive and heterogeneous CQR effects do not imply that education has a stronger effect for the highest wage earners, and this then contributes to increase inequality. Instead it means that it has a stronger effect for the conditionally rich, that is, after controlling for all other covariates. The advantage of the UQR approach is that it allows study the effects directly on the distribution of income. The UQR results show also a heterogeneous behavior of the returns to schooling along the wage distribution, but it is even more pronounced. Returns to schooling range from 1.18% to 16.17% for the Golden Triangle, and from 4.19% to 8.99% in the Pacific region. With the results from UQR now it can be said that the returns to schooling are larger for those individuals located in the upper part of the wage distribution, those with the highest wages. As with the CQR, returns to schooling are higher in the Pacific region at lower quantiles, compare to those in Golden Triangle. However, in contrast to what was found with CQR, at the middle part of the distribution, the returns are considerably higher in the Golden Triangle.

The informal pay penalty decreases sharply from the lower quantile to the middle quantile and is statistically significant mainly for the lower quantiles. In the case of the Golden Triangle, and according to the CQR results, the pay penalty is of around 16.62% in the lowest quantile and 7.26% in the upper quantile. In Pacific region informal workers faced a penalty of 29.59% in the lowest quantile and 17.37% in the highest. The informal effect at higher quantiles in the case of the UQR turn to be positive for some regions (e.g. Atlantic, Oriental and Central), pointing towards the existence of a premium for informal workers, although such positive coefficients lack statistical significance. Informality affects negatively mostly those individuals positioned at the lower part of conditional and unconditional wage distribution. The decrease in the pay penalty of informality means that a 1 percentage increase in informal jobs decreases wages more at the bottom than at the top of the wage distribution. In other words a rise in informal jobs will increase wage inequality in all the Colombians regions.

These estimates confirm the positive effect of education on wages and an increasing effect at higher quantiles of the wage distribution. There is substantial regional variability in the returns to schooling. Furthermore, they suggest that difference in returns to years of education may be an important factor explaining wage differentials across regions. On the other hand, workers face different informal pay penalties throughout the territory and it affects mostly individuals at the lower part of the wage distribution. Therefore its contribution in explaining regional wage gaps may be limited to this part.

The evidence presented so far confirms that regions not only differed in the endowment of earning relevant characteristics, such as education and the incidence of informality, but also shows sizeable regional variability in the returns to these characteristics. The next section assesses the contribution of this variability in characteristics and returns to the wage gap across regions.

4.4.3 Decomposition of regional wage gaps

The decomposition of regional wage differentials in Colombia is analyzed by considering the difference between Golden Triangle, the region with the highest level of wages, and other regions. Estimated regional wage differentials for each region relative to Golden Triangle for the mean and for the selected quantiles are reported in the first row of Table 4.5. It also contains the global decomposition, in which wage gaps are decomposed in two terms, one that accounts for the contribution attributable to difference in observable characteristics (labeled Total explained by characteristics) and another that corresponds to differences in the wage structure (labeled Total wage structure). Both of these two components can in turn be decomposed in the specific contribution of each factor that determine wages, by using the detail decomposition. Given the main interest in this chapter, the details of the specific contribution of education and informality are presented in Table 4.5, while the contributions of the rest of control variables have been grouped in the term labeled *rest*. In addition, results from the decomposition with and without the reweighting are presented in panels A and B respectively.

Wage differentials between regions, calculated at the mean, are all statistically significant. The highest wage gap is found between the Pacific region and the Golden Triangle, 36%, while the lowest one is that of the Golden Triangle and Atlantic, 9%. Results from the global decomposition without reweighting (Panel A) indicate that the contributions of coefficients are larger than that of characteristics for most of the regions, except for Oriental region. In the case of Atlantic, difference in characteristics pushes down the wage gap, as this region is more endowed than Golden Triangle. However difference in coefficients enlarged the wage gap, meaning that workers characteristics in Golden Triangle are better rewarded than in Atlantic region.

In all regions, except for Atlantic, the specific contribution of education indicates that a considerable part of the wage differential between regions is explained by the fact that the Golden Triangle has a more educated workforce. Golden Triangle also displays the highest returns to schooling, which is reflected in the positive effect in the wage structure. Meanwhile the differences in the incidence of informality across regions suggest that a more equal distribution of informality may reduce the wage gap between regions. In contrast, the difference in the informal pay penalty does not contribute to drive regional wage gaps.

As already discussed, wage differentials at the mean may hide important information of the wage gap across the wage distribution. Table 4.5 also shows regional wage differentials for each region relative to Golden Triangle at different quantiles. The quantile approach reveals that, for Oriental and Pacific regions, the wage gap along the wage distribution has a non-monotonic behavior. This behavior is different to what has been described for developed countries. Motellón, López-Bazo and El-Attar (2011) found an increasing wage differential across the wage distribution for Spain and Pereira and Gallego (2013) found the same pattern for Portugal.

Regional wage gaps and the decomposition analysis at selected quantiles employing the method in Firpo, Fortin and Lemieux (2009) are also reported in Table 4.5. For most of the regions and for most of the quantiles differences in coefficients are the dominant effect explaining regional wage gaps. However, Oriental once again stands as the region in which difference in characteristics represents the most part of the wage differential. The specific contribution of education at lower quantiles is not what is driving regional wage differentials. If any, in some cases education pushes down wage differentials at lower quantiles. For example, in the case of Pacific at the 25th quantile difference in returns to schooling reduce the wage gap. However at the middle and at higher quantiles of the wage distribution education plays an

important role and a large part of wage differentials is due to difference in the returns to education. As expected, informality and specifically its incidence only affect regional wage gaps at lower quantiles. Informality represents around 50% of the wage gap at the 25th quantile in Pacific, meaning that reducing informality in this region will help to reduce the wage gap considerably.

With respect to the constant, it is only important in the case of the Oriental region. The constant corresponds to the unexplained part, not accounted by covariates. For the other regions is not statistically significant.

Table 4.5 displays the decomposition with reweighting. Concerning the reweighting decomposition, one can see that the results change slightly for most regions. However in the case of the Pacific the reweighting decomposition points to a greater contribution of the characteristics component to the wage gap and less to the wage structure, though it remains considerable for the lowest quantile. The specification error is for some regions and for certain quantiles statistically significant and its value is not negligible. As for the reweighting errors, they are quite small for most quantiles and sometimes significant at 5% level. Nevertheless the conclusions derived from the decomposition without reweighting remain fairly the same for most of the regions.

These results lead to the conclusion that policies aiming at reducing human capital differences among regions will help to decrease regional wage gaps, especially at the higher parts of the wage distribution. However, equalizing years of education of workers across regions would not be enough to reduce regional wage differences due to the sizeable differences in returns to years of education at higher quantiles. Similar results have been found in previews studies, albeit in a context of developed countries. Meanwhile policies that point towards the reduction of informality will help to lower

regional wage gaps at the lower part of the wage distribution particularly for those regions with sizable informality.

4.5 Regional formal and informal wage gaps

The above results were done jointly for formal and informal workers, thus assuming that returns to education and the effects of other relevant characteristics that determine wages were the same for both type of workers. However, the existence of institutional arraignment or different wage structures may affect the way formal and informal workers are rewarded, and therefore the prices that they perceived for their characteristics. For example, it is well known that the minimum wage is binding in the formal sector, meaning that a large proportion of formal workers earn a minimum wage, in contrast a large proportion of informal workers are paid a wage inferior to the minimum wage. This adds to the fact that the share of workers in the informal sector varies largely across Colombian regions. Thus, grouping formal and informal workers together may give misleading information about the origin of regional wage disparities.

As shown in Figure 4.3, once the density of hourly wages is computed for formal and informal workers separately, the regional differences are less marked within each of these two groups of workers. As a matter of fact, Pacific region whose density distribution of hourly wages had a very dissimilar behavior compared to other regions in the total sample, once formal and informal workers are treated separately, its behavior is more alike, especially in the case of formal workers. Table 4.6 provides a description of hourly wages for formal and informal workers separately and for the five regions, similar to Table 4.2. Undoubtedly, average hourly wages are different between regions, even after splitting the sample into formal and informal workers. However wage gaps are reduced considerably. By comparing formal workers from Pacific region to formal workers of Golden Triangle now the average of gross

hourly wages of the Pacific is 95% of that paid in Golden Triangle. When considering only informal workers, the wage gap of Pacific against Golden Triangle is also reduced, although to a lesser extent than when comparing formal workers. The last columns of Table 4.6 report gross hourly wages at the selected percentiles. The wage gap for formal workers behaves in a different way along the wage distribution for each of the regions, while a non-monotonic behavior throughout the wage distribution is present for informal workers. Since the magnitude and the behavior of regional wage gaps of formal workers are different to those of informal workers, then treating formal and informal workers separately will complement the analysis and will give a more complete understanding of regional wage gaps in a labor market characterized by a high degree of informality. In doing so this section will present the same analysis done so far, but differentiating formal and informal workers. However, the focus is not to compare formal and informal workers, but to compare formal workers across regions and separately doing the same for informal workers. While comparing formal workers to their informal counterparts across regions is of interest, is out of the scope of this study (Garcia, 2014 examines the heterogeneity of the formal/informal wage gap at the regional level in Colombia). Moreover, the selectivity bias associated with non-observable characteristics that could simultaneously affect wages and the sector in which the individuals are currently working is less likely to affect the results when comparing formal (informal) workers of one region with formal (informal) workers of other regions.

Table 4.7 reports the results concerning the estimates of the Mincer wage equations by OLS and by quantile regressions (conditional and unconditional) for the selected quantiles for the five regions and for formal and informal workers separately. The discussion that follows this set of results is done taking as a point of reference the results in Table 4.4 when formal and informal workers were treated jointly with the aim of highlighting the

importance of this subsequent analysis. The description of the results will focus only on the returns to education. Looking at the results found for formal workers it is observed that returns to education for these types of workers differ across regions but to a lesser extend than those obtained previously. Results from quantile regression (conditional and unconditional) show that returns to education for formal workers increase along the wage distribution and for specific quantiles some differences between regions exist, but again these differences are lower than those found in Table 4.4.

Turning now to the results for informal workers it is visible that returns to education differ considerably across regions for this type of workers. The OLS estimates show that informal workers of the Atlantic and Pacific regions have the highest returns, around 5%, while Oriental and, surprisingly, Golden Triangle display the lowest ones, around 3%. The returns to education for informal workers increase along the wage distribution in some cases, such as in Central and Oriental regions, and for other regions they have a non-monotonic behavior. Clearly the results for informal workers differ considerably to those found in Table 4.4. Moreover they suggest that the value of additional education is quite constraint in the informal sector, as more education not necessarily means higher wages.

From these findings it is clear that grouping formal and informal workers does not reveal the complete picture and may produce only incomplete conclusions. There are reasons to suspect that the decomposition analysis might also give new information if it is done for formal and informal workers separately. Table 4.8 and Table 4.9 display the results of the decomposition exercise for formal and informal workers respectively, similar to the results presented in Table 4.5 for the entire sample of workers. Results from the global decomposition, for formal and informal workers, show that for the Atlantic region the results are fairly the same, but for the rest of the regions the results of the decomposition provide new information. First, it is

important to notice that for all regions the characteristics component reduces considerably its contribution. This may be the result of comparing more homogenous workers across regions, especially in the case of formal workers who share similar worker and firms' characteristics. For the Pacific region it now turns that formal workers are better endowed that formal workers in the Golden Triangle, and thus the characteristics component reduce the wage gap between these two regions. In the case of Central region the component corresponding to differences in characteristics is not statistically significant neither for formal nor for informal workers. However it remain true that the characteristics of workers in Golden Triangle are better-rewarded compared to other regions.

The detailed decomposition, and particularly the contribution of education, also varies considerably once the analysis is done for formal and informal workers independently. In the case of formal workers, the Atlantic region is endowed with workers with more education than those formal workers in the Golden Triangle. In other words, the wage gap between Atlantic and Golden Triangle would have been lower if they had differed only in the average years of education of their workers. In the case of the Oriental region, the difference in the endowment of human capital contributed to widen the wage gap with Golden Triangle. As for the Pacific region, the results for the decomposition indicate that the differences in the endowment of human capital are not driving the wage gap. If any, differences in education seems to be lowering the wage differentials across the quantiles, though in the case of the reweighted decomposition human capital differences are not statistically significant. For the central region the difference in education is not statistically significant. Regarding the difference in the returns to education, the results for formal workers confirm that the difference in returns to human capital contributed to increase the wage gap.

These results revels that some of the conclusions derived from the preview analysis that treated formal and informal workers jointly are partially correct. For instance the belief that the Golden Triangle is the region with the largest endowed workforce is not completely accurate. Moreover the distribution of education is generating an equalizing effect of wages across some regions. While the returns to education continue to be a source of wage inequality across Colombian territories.

4.6 Conclusions

Results from micro-data for Colombia confirmed the existence of differences not only in average regional wages but also across the wage distribution. This study used the decomposition approach proposed by Firpo, Fortin and Lemieux (2009) to estimate the contributions of regional differences in characteristics and of regional differences in the wage structures to the observed regional wage gaps. This methodology has the advantage that allows estimating the contribution of each characteristic along the entire wage distribution. Given that Colombian regions are characterized by significant differences in the education of their workforce and in the incidence of informality, the contribution of both of these two factors to the regional wage gaps are closely examined.

The results of the decomposition for Colombia show that for most of the regions and for most of the quantiles differences in the wage structures are the dominant factor explaining regional wage gaps. Meaning that workers with similar characteristics received different wages depending on the region in which they are located. At the middle and especially at higher quantiles of the wage distribution education plays an important role and a large part of wage differentials is due to differences in the returns to education. Informality and specifically its incidence only affect regional wage gaps at lower quantiles. Therefore policies that points towards the reduction of informality will help to

lower regional wage gaps at the lower part of the wage distribution particularly for those regions with sizable informality.

This study has shown the importance of examining regional wage gaps separately for formal and informal workers since, in addition to the regional disparities in the incidence of informality, it has been proved that the wage structure differ between the two sectors. Accordingly, the reasons behind regional wage disparities when distinguishing between workers of the two sectors may deviate from those found when they are grouped together. Wage gaps are reduced considerably once formal workers are compared between regions, particularly for those regions with a high incidence of informality. Suggesting that formalization of employment, aside from the well-known implications of higher wages and social security coverage, may also help reducing disparities across regions. Moreover, if regional labor markets are segmented and formal and informal jobs are characterized by different mechanisms of functioning and adjustment, the proposed policy may not be unique for each of these two segments.

As in past studies of this nature, it remains to be explained why the difference in the returns to education across regions is persistent. We hypothesize that such a difference in returns is related to economies of scale and agglomeration economies; however further research is clearly required on this matter for a better understanding of regional wage differentials.

Table 4.1. Hourly wage, informality and human capital variables for the thirteen largest metropolitan of Colombia

	Number of Observations	Nominal Gross Hourly wage (pesos)	Adjusted Hourly wage (pesos)	Schooling (years)	Informality (%)
By metropolitan area					
Barranquilla	1037	3663.16	3510.73	11.31	35.29
		(2947.25)	(2824.61)	(3.45)	
Cartagena	809	3760.54	3605.99	11.74	22.00
		(2518.59)	(2415.08)	(3.44)	
Monteria	759	3650.30	3493.12	11.26	36.89
		(3218.13)	(3079.56)	(3.59)	
Cucuta	754	2825.23	2634.22	9.39	59.15
		(1837.99)	(1713.73)	(4.07)	
Bucaramanga	988	3662.94	3442.25	10.65	31.88
		(2562.04)	(2407.68)	(3.87)	
Villavicencio	862	3306.05	3141.81	10.11	43.85
		(2464.41)	(2341.98)	(3.48)	
Manizales	1109	3506.84	3402.62	11.19	20.83
		(2680.53)	(2600.87)	(3.74)	
Pereira	1014	3351.98	3230.37	10.24	28.60
		(2547.55)	(2455.12)	(3.89)	
Ibague	869	3678.27	3501.31	11.06	36.02
		(2913.20)	(2773.05)	(3.73)	
Pasto	733	2981.61	2885.20	10.53	49.39
		(2668.21)	(2581.93)	(4.14)	
Medellin	1913	3903.84	3718.43	10.96	18.98
		(2904.72)	(2766.76)	(3.76)	
Bogota	1754	4305.70	4132.05	11.33	23.95
		(3566.44)	(3422.61)	(3.96)	
Cali	1195	3872.52	3745.43	10.68	28.62
		(3147.60)	(3044.30)	(3.83)	
Colombia	13796	3662.54	3504.48	10.86	31.05
		(2894.79)	(2773.67)	(3.82)	

Notes: Sample means (standard deviation are shown for continuous variables).

Table 4.2. Descriptive of *adjusted* hourly wages in the five regions of Colombia

				Percentiles				
	Average	Std. Dev. of Logs	Gini	10%	25%	50%	75%	90%
Atlantic	3535.18	0.57	0.33	1631.12	2395.67	2617.42	3727.07	6496.88
Oriental	3108.82	0.54	0.31	1478.28	2000.76	2489.83	3321.36	5188.98
Central	3372.9	0.54	0.32	1635.19	2144.57	2467.86	3489.06	6015.41
Pacific	2885.19	0.69	0.39	940.79	1458.48	2325.62	3010.51	5644.71
Golden Triangle	3874.31	0.57	0.34	1874.24	2384.57	2778.14	4167.22	7165.54
Atlantic vs. Golden	0.09	-	-	0.13	0.00	0.06	0.11	0.09
Oriental vs. Golden	0.20	-	-	0.21	0.16	0.10	0.20	0.28
Central vs. Golden	0.13	-	-	0.13	0.10	0.11	0.16	0.16
Pacific vs. Golden	0.26	-	-	0.50	0.39	0.16	0.28	0.21

Notes: Sample means. Wage gap = (golden - region$_i$)/region$_i$.

Table 4.3. Descriptive of observable worker and firm characteristics

	Atlantic	Oriental	Central	Pacific	Golden Triangle
Adjusted Hourly Wage	3535.18	3108.82	3372.9	2885.19	3874.28
Informal	0.32	0.44	0.28	0.49	0.23
Worker's characteristics					
Schooling (years)	11.43	10.10	10.83	10.53	11.03
Experience (years)	18.02	17.09	18.55	17.99	18.05
Tenure (months)	53.91	36.92	48.57	44.74	50.21
Women	0.39	0.43	0.42	0.43	0.45
Married	0.60	0.51	0.49	0.52	0.51
Head of household	0.43	0.40	0.43	0.43	0.44
Type of contract					
No-contract	0.25	0.44	0.26	0.43	0.23
Temporary	0.21	0.21	0.24	0.28	0.24
Permanent	0.54	0.36	0.50	0.29	0.52
Firm size					
Micro	0.27	0.44	0.32	0.50	0.28
Small	0.21	0.21	0.18	0.16	0.20
Medium	0.06	0.05	0.05	0.02	0.07
Large	0.46	0.31	0.44	0.32	0.45
Sector					
Mining, electricity, gas and water	0.04	0.04	0.03	0.02	0.02
Industry	0.21	0.19	0.24	0.16	0.26
Construction	0.05	0.11	0.08	0.07	0.06
Sales, Hotels and Restaurants	0.29	0.34	0.26	0.38	0.27
Transportation	0.10	0.08	0.10	0.08	0.07
Financial Intermediation	0.11	0.09	0.11	0.07	0.15
Social Services	0.20	0.16	0.18	0.22	0.17
Observations	*2605*	*2604*	*2992*	*733*	*4862*

Notes: Sample means.

Table 4.4. Estimations of returns to education and informality for five regions of Colombia - OLS and quantiles estimates (conditional and unconditional)

		CQR			UQR		
	OLS	25	50	75	25	50	75
Years of education							
Atlantic	0.0826**	0.0553**	0.0697**	0.0873**	0.0087**	0.0435**	0.1319**
	[0.0028]	[0.0020]	[0.0029]	[0.0035]	[0.0012]	[0.0025]	[0.0056]
Oriental	0.0557**	0.0353**	0.0440**	0.0557**	0.0215**	0.0253**	0.0740**
	[0.0027]	[0.0024]	[0.0025]	[0.0035]	[0.0036]	[0.0022]	[0.0046]
Central	0.0752**	0.0412**	0.0569**	0.0779**	0.0214**	0.0306**	0.1148**
	[0.0024]	[0.0016]	[0.0023]	[0.0034]	[0.0024]	[0.0016]	[0.0048]
Pacific	0.0682**	0.0516**	0.0659**	0.0729**	0.0419**	0.0288**	0.0899**
	[0.0050]	[0.0051]	[0.0053]	[0.0083]	[0.0099]	[0.0051]	[0.0079]
Golden Triangle	0.0814**	0.0462**	0.0674**	0.0899**	0.0118**	0.0519**	0.1617**
	[0.0020]	[0.0014]	[0.0017]	[0.0032]	[0.0011]	[0.0019]	[0.0047]
Colombia	0.0742**	0.0460**	0.0597**	0.0778**	0.0139**	0.0374**	0.1254**
	[0.0012]	[0.0009]	[0.0011]	[0.0020]	[0.0009]	[0.0010]	[0.0024]
Informality							
Atlantic	-0.1023**	-0.1691**	-0.0435+	-0.0475+	-0.1137**	-0.0874**	-0.0472
	[0.0257]	[0.0192]	[0.0264]	[0.0265]	[0.0138]	[0.0258]	[0.0525]
Oriental	-0.0991**	-0.1341**	-0.0516*	-0.0599*	-0.2710**	-0.0810**	0.0123
	[0.0257]	[0.0224]	[0.0234]	[0.0302]	[0.0355]	[0.0231]	[0.0445]
Central	-0.0951**	-0.1704**	-0.0515*	0.0159	-0.2389**	-0.0572**	0.0414
	[0.0274]	[0.0183]	[0.0263]	[0.0341]	[0.0326]	[0.0215]	[0.0493]
Pacific	-0.2680**	-0.2959**	-0.2422**	-0.1737+	-0.3085*	-0.3499**	-0.2939**
	[0.0558]	[0.0573]	[0.0595]	[0.0893]	[0.1200]	[0.0642]	[0.0868]
Golden Triangle	-0.1356**	-0.1662**	-0.1091**	-0.0726*	-0.1473**	-0.0470+	-0.0215
	[0.0227]	[0.0169]	[0.0195]	[0.0298]	[0.0147]	[0.0249]	[0.0487]
Colombia	-0.1430**	-0.1927**	-0.0891**	-0.0856**	-0.1881**	-0.0917**	-0.0471+
	[0.0125]	[0.0096]	[0.0116]	[0.0186]	[0.0109]	[0.0118]	[0.0242]

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

Table 4.5. Regional wage gap decomposition

Atlantic	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.087**	**	**0.006**		**0.068**	**	**0.114**	**	**0.087**	**	**0.006**		**0.068**	**	**0.114**	**
Composition Effect attributable to																
Education	-0.033	**	-0.005	**	-0.021	**	-0.065	**	-0.020	**	-0.002	*	-0.009	*	-0.033	*
Informality	0.012	**	0.013	**	0.004	+	0.002		0.012	**	0.013	**	0.007	**	-0.006	
Rest	-0.027	**	-0.011	**	-0.024	**	-0.050	**	-0.023	**	-0.008	**	-0.013	**	-0.017	+
Error									0.010		-0.002		-0.007		0.027	
Total explained by characteristics	**-0.049**	**	**-0.003**		**-0.041**		**-0.114**	**	**-0.021**	**	**0.002**		**-0.022**	*	**-0.030**	
Wage structure effects attributable to																
Education	-0.014		0.035	+	0.096	*	0.340	**	0.052		0.048	**	0.190	**	0.378	**
Informality	-0.011		-0.011	*	0.013		0.008		0.000		0.001		0.009		-0.022	
Rest	0.110		-0.093	*	0.074		0.003		0.137		0.004		0.0345		-0.200	
Constant	0.051		0.077	*	-0.073		-0.124		-0.063		-0.044		-0.130		0.025	
Error									-0.018	+	-0.005		-0.013	*	-0.037	*
Total wage structure	**0.136**	**	**0.009**	+	**0.110**	**	**0.228**	**	**0.109**	**	**0.004**		**0.091**	**	**0.144**	**
Oriental	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.190**	**	**0.187**	**	**0.118**	**	**0.238**	**	**0.190**	**	**0.187**	**	**0.118**	**	**0.238**	**
Composition Effect attributable to																
Education	0.075	**	0.011	**	0.048	**	0.149	**	0.067	**	0.014	**	0.045	**	0.113	**
Informality	0.028	**	0.030	**	0.010	*	0.004		0.017	**	0.035	**	0.004		-0.018	
Rest	0.063	**	0.044	**	0.053	**	0.075	**	0.090	**	0.057	**	0.075	**	0.114	**
Error									-0.004		0.058	**	-0.008		0.013	
Total explained by characteristics	**0.166**	**	**0.086**	**	**0.111**	**	**0.228**	**	**0.171**	**	**0.165**	**	**0.115**	**	**0.221**	**
Wage structure effects attributable to																
Education	0.260	**	-0.098	**	0.270	**	0.885	**	0.100	**	-0.037	**	0.046		0.450	**
Informality	-0.016		0.054	**	0.015		-0.015		-0.012		0.005		-0.006		-0.026	
Rest	0.041		0.389	**	0.0727		-0.137		0.139		0.174	**	0.118		-0.211	
Constant	-0.260	**	-0.244	*	-0.350	**	-0.724	**	-0.203	+	-0.118		-0.151		-0.188	
Error									-0.004		0.024		0.007		-0.009	
Total wage structure	**0.025**		**0.101**	**	**0.007**		**0.010**		**0.019**		**0.022**		**0.003**		**0.017**	

Notes: + p<0.1, * p<0.05, ** p<0.01.

Table 4.5 continue

Central	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.119**	**	**0.111**	**	**0.127**	**	**0.189**	**	**0.119**	**	**0.111**	**	**0.127**	**	**0.189**	**
Composition Effect attributable to																
Education	0.016	*	0.002	*	0.010	*	0.031	*	0.021	**	0.005	**	0.010	**	0.035	**
Informality	0.006	**	0.007	**	0.002	+	0.001		0.005	**	0.010	**	0.003	*	-0.004	
Rest	0.007		0.009	**	0.007		-0.001		0.007		0.011	*	0.006		0.007	
Error									0.003		0.017	*	0.021	**	0.012	
Total explained by characteristics	**0.029**	**	**0.018**	**	**0.019**	**	**0.031**	**	**0.036**	**	**0.043**	**	**0.040**	**	**0.051**	**
Wage structure effects attributable to																
Education	0.067	*	-0.104	**	0.232	**	0.507	**	0.029		-0.061	**	0.152	**	0.339	**
Informality	-0.011		0.026	**	0.003		-0.018		-0.005		0.014	*	0.004		-0.025	
Rest	0.016		0.321	**	-0.029		-0.077		0.025		0.222	**	0.005		-0.069	
Constant	0.018		-0.149	+	-0.098		-0.255		0.042		-0.104		-0.069		-0.093	
Error									-0.007		-0.002		-0.005		-0.014	
Total wage structure	**0.090**	**	**0.093**	**	**0.108**	**	**0.157**	**	**0.083**	**	**0.069**	**	**0.087**	**	**0.138**	**
Pacific	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.362**	**	**0.499**	**	**0.180**	**	**0.334**	**	**0.362**	**	**0.499**	**	**0.180**	**	**0.334**	**
Composition Effect attributable to																
Education	0.040	**	0.006	**	0.026	**	0.080	**	0.065	**	0.023	**	0.040	**	0.105	**
Informality	0.036	**	0.039	**	0.012	+	0.006		0.066	**	0.165	**	0.049	**	0.022	
Rest	0.033	**	0.018	**	0.025	**	0.042	*	0.078	**	0.098	**	0.049	*	0.075	*
Error									-0.002		0.018		-0.091	**	0.080	+
Total explained by characteristics	**0.108**	**	**0.062**	**	**0.063**	**	**0.127**	**	**0.207**	**	**0.303**	**	**0.047**	+	**0.283**	**
Wage structure effects attributable to																
Education	0.139	**	-0.317	**	0.243	**	0.755	**	0.068		-0.169	*	0.054		0.443	**
Informality	0.065	*	0.080		0.150	**	0.135	**	0.028	+	0.117	**	0.034	*	0.016	
Rest	-0.140		-0.121		0.011		-0.401		-0.093		0.319		-0.099		-0.231	
Constant	0.189		0.796	*	-0.287		-0.283		0.179		-0.063		0.164		-0.123	
Error									-0.028	+	-0.008		-0.021	+	-0.054	+
Total wage structure	**0.253**	**	**0.437**	**	**0.117**	**	**0.207**	**	**0.155**	**	**0.196**	**	**0.132**		**0.051**	

Notes: + p<0.1, * p<0.05, ** p<0.01.

Table 4.6. Descriptive of hourly wages for formal and informal workers

Formal

				Percentiles				
	Average	Std. Dev. of Logs	Gini	10%	25%	50%	75%	90%
Atlantic	4070.65	0.509	0.31	2395.67	2400.59	2888.48	4195.21	7442.87
Oriental	3805.30	0.508	0.30	2039.23	2352.64	2838.29	4111.41	6834.07
Central	3812.80	0.501	0.31	2078.56	2412.64	2793.18	3967.82	6791.97
Pacific	4101.96	0.553	0.34	1966.44	2422.52	2822.36	4515.77	8429.44
Golden Triangle	4292.18	0.548	0.34	2289.19	2402.51	2985.64	4665.06	8260.35
Atlantic vs. Golden	0.05	-	-	-0.04	0.00	0.03	0.11	0.11
Oriental vs. Golden	0.13	-	-	0.12	0.02	0.05	0.13	0.21
Central vs. Golden	0.13	-	-	0.10	0.00	0.07	0.18	0.22
Pacific vs. Golden	0.05	-	-	0.16	-0.01	0.06	0.03	-0.02

Informal

				Percentiles				
	Average	Std. Dev. of Logs	Gini	10%	25%	50%	75%	90%
Atlantic	2377.82	0.53	0.29	1063.266	1594.9	2232.859	2608.945	3577.982
Oriental	2213.00	0.44	0.23	1208.654	1582.239	2105.398	2610.694	3289.126
Central	2234.65	0.47	0.25	1181.891	1572.214	2056.551	2498.527	3223.331
Pacific	1638.18	0.51	0.28	711.6368	1023.022	1477.888	2037.228	2533.064
Golden Triangle	2486.12	0.48	0.26	1267.638	1751.266	2256.756	2799.037	3776.234
Atlantic vs. Golden	0.05	-	-	0.19	0.10	0.01	0.07	0.05
Oriental vs. Golden	0.12	-	-	0.05	0.11	0.07	0.07	0.13
Central vs. Golden	0.11	-	-	0.07	0.11	0.09	0.11	0.15
Pacific vs. Golden	0.52	-	-	0.78	0.71	0.35	0.27	0.33

Notes: Sample means. Wage gap = (golden - region $_i$)/region $_i$.

Table 4.7. Estimations of returns to education for five regions of Colombia for formal and informal workers - OLS and quantiles estimates (conditional and unconditional)

		CQR			UQR		
	OLS	25	50	75	25	50	75
Formal							
Atlantic	0.0955**	0.0554**	0.0845**	0.1041**	0.0218**	0.0769**	0.1693**
	[0.0034]	[0.0020]	[0.0031]	[0.0051]	[0.0022]	[0.0037]	[0.0079]
Oriental	0.0740**	0.0371**	0.0623**	0.0845**	0.0157**	0.0509**	0.1206**
	[0.0036]	[0.0026]	[0.0033]	[0.0070]	[0.0025]	[0.0037]	[0.0074]
Central	0.0879**	0.0371**	0.0753**	0.0952**	0.0127**	0.0644**	0.1803**
	[0.0028]	[0.0013]	[0.0029]	[0.0061]	[0.0016]	[0.0032]	[0.0069]
Pacific	0.0824**	0.0576**	0.0799**	0.0980**	0.0163**	0.0718**	0.1605**
	[0.0070]	[0.0039]	[0.0052]	[0.0148]	[0.0053]	[0.0083]	[0.0170]
Golden Triangle	0.0950**	0.0515**	0.0808**	0.1062**	0.0182**	0.0777**	0.1699**
	[0.0022]	[0.0015]	[0.0025]	[0.0042]	[0.0014]	[0.0025]	[0.0049]
Colombia	0.0890**	0.0449**	0.0756**	0.0975**	0.0159**	0.0707**	0.1728**
	[0.0014]	[0.0007]	[0.0018]	[0.0027]	[0.0008]	[0.0016]	[0.0034]
Informal							
Atlantic	0.0551**	0.0480**	0.0409**	0.0429**	0.0340**	0.0231**	0.0330**
	[0.0051]	[0.0083]	[0.0066]	[0.0052]	[0.0077]	[0.0042]	[0.0050]
Oriental	0.0289**	0.0258**	0.0266**	0.0251**	0.0178**	0.0284**	0.0274**
	[0.0041]	[0.0059]	[0.0035]	[0.0034]	[0.0059]	[0.0045]	[0.0045]
Central	0.0486**	0.0374**	0.0349**	0.0401**	0.0288**	0.0295**	0.0363**
	[0.0046]	[0.0054]	[0.0040]	[0.0044]	[0.0057]	[0.0044]	[0.0045]
Pacific	0.0508**	0.0497**	0.0543**	0.0451**	0.0290**	0.0522**	0.0471**
	[0.0069]	[0.0097]	[0.0102]	[0.0089]	[0.0109]	[0.0098]	[0.0098]
Golden Triangle	0.0346**	0.0277**	0.0226**	0.0291**	0.0186**	0.0183**	0.0250**
	[0.0044]	[0.0057]	[0.0040]	[0.0046]	[0.0052]	[0.0038]	[0.0048]
Colombia	0.0415**	0.0344**	0.0322**	0.0324**	0.0255**	0.0287**	0.0309**
	[0.0022]	[0.0030]	[0.0019]	[0.0020]	[0.0032]	[0.0024]	[0.0023]

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

Table 4.8. Regional wage gap decomposition for formal workers

Atlantic	**A. Without reweighting**								**B. With reweighting**							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.032**	*	**0.015**	+	**0.032**	*	**0.099**	**	**0.032**	**	**0.015**	+	**0.032**	*	**0.099**	**
Composition Effect attributable to																
Education	-0.055	**	-0.011	**	-0.045	**	-0.098	**	-0.040	**	-0.004	**	-0.030	**	-0.067	**
Rest	-0.043	**	-0.022	**	-0.041	**	-0.062	**	-0.037	**	-0.016	**	-0.027	**	-0.033	**
Error									0.011		0.054	**	0.044	**	-0.006	
Total explained by characteristics	**-0.098**	**	**-0.033**	**	**-0.086**	**	**-0.160**	**	**-0.066**	**	**0.034**	**	**-0.014**		**-0.106**	**
Wage structure effects attributable to																
Education	-0.007		-0.044		0.010		0.007		0.161	**	0.114	**	0.201	**	0.398	**
Rest	0.070		0.074		0.065		0.008		0.033		0.119	+	0.211		-0.272	
Constant	0.067		0.018		0.043		0.244		-0.077		-0.247	**	-0.351	*	0.116	
Error									-0.019	+	-0.005		-0.015		-0.037	*
Total wage structure	**0.130**	**	**0.048**	**	**0.118**	**	**0.259**	**	**0.098**	**	**-0.020**	**	**0.046**	**	**0.205**	**
Oriental	**A. Without reweighting**								**B. With reweighting**							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.009**	**	**0.010**		**0.053**	**	**0.130**	**	**0.090**	**	**0.010**		**0.053**	**	**0.130**	**
Composition Effect attributable to																
Education	0.024	*	0.005	*	0.020	*	0.043	*	0.019	+	0.006	+	0.017	+	0.034	+
Rest	0.035	**	0.015	**	0.033	**	0.044	**	0.050	**	0.021	**	0.045	**	0.076	**
Error											-0.035	**	0.002		-0.007	
Total explained by characteristics	**0.059**	**	**0.020**	**	**0.053**	**	**0.087**	**	**0.069**	**	**-0.007**		**0.063**	**	**0.104**	**
Wage structure effects attributable to																
Education	0.239	**	0.028		0.305	**	0.561	**	0.152	**	-0.074	+	0.084		0.289	**
Rest	0.035		0.207	**	0.093		-0.217		0.092		0.192	+	0.080		-0.129	
Constant	-0.243	+	-0.244	**	-0.398	**	-0.301		-0.219		-0.098		-0.170		-0.124	
Error									-0.005		-0.003		-0.005		-0.009	
Total wage structure	**0.031**	*	**-0.010**		**0.000**		**0.043**	+	**0.021**		**0.017**		**-0.010**		**0.027**	

Notes: + p<0.1, * p<0.05, ** p<0.01.

Table 4.8 continue

Central	A. Without reweighting				B. With reweighting			
		Quantiles				Quantiles		
	OLS	25	50	75	OLS	25	50	75
Overall wage gap	**0.093 ****	**0.003**	**0.069 ****	**0.169 ****	**0.093 ****	**0.003**	**0.069 ****	**0.169 ****
Composition Effect attributable to								
Education	-0.006	-0.001	-0.005	-0.011	-0.001	0.000	-0.001	-0.002
Rest	0.006	0.004	0.005	0.006	0.008	0.006	0.008	0.007
Error					0.007	0.001	0.007	0.016
Total explained by characteristics	**0.000**	**0.003**	**0.000**	**-0.005**	**0.014**	**0.007**	**0.014**	**0.021**
Wage structure effects attributable to								
Education	0.083 *	0.064 *	0.156 **	-0.122	0.094 *	0.059 *	0.114 *	-0.162
Rest	0.100	0.137 *	0.245 +	0.208	0.067	0.120 +	0.226 +	0.188
Constant	-0.089	-0.200 **	-0.332 *	0.089	-0.074	-0.182 *	-0.278 +	0.136
Error					-0.007	-0.002	-0.006	-0.012
Total wage structure	**0.094 ****	**0.000**	**0.069 ****	**0.175 ****	**0.079 ****	**-0.004**	**0.055 ****	**0.149 ****

Pacific	A. Without reweighting				B. With reweighting			
		Quantiles				Quantiles		
	OLS	25	50	75	OLS	25	50	75
Overall wage gap	**0.043**	**-0.025**	**0.055 +**	**0.018**	**0.043**	**-0.025**	**0.055 +**	**0.018 0**
Composition Effect attributable to								
Education	-0.058 **	-0.011 **	-0.048 **	-0.104 **	-0.018	-0.005	-0.020	-0.024
Rest	0.001	-0.012 +	-0.005	0.014	-0.033	-0.011	0.028	-0.006
Error					-0.008	0.004	-0.026	-0.038
Total explained by characteristics	**-0.057 ***	**-0.023 ****	**-0.053 ****	**-0.089 ***	**-0.060**	**-0.012**	**-0.018**	**-0.069**
Wage structure effects attributable to								
Education	0.154 +	0.022	0.073	0.114	0.231 **	-0.020	-0.059	0.820 **
Rest	0.034	0.405 *	-0.027	-0.099	-0.034	0.394 *	-0.065	-0.436
Constant	-0.088	-0.429 *	0.061	0.092	-0.070	-0.382 *	0.218	-0.250
Error					-0.026	-0.006	-0.021	-0.048
Total wage structure	**0.100 ****	**-0.002**	**0.108 ****	**0.107 ***	**0.102 ****	**-0.013**	**0.073 ***	**0.087 ***

Notes: + p<0.1, * p<0.05, ** p<0.01.

Table 4.9. Regional wage gap decomposition for informal workers

Atlantic	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.077**	**	**0.094**	**	**0.011**		**0.083**	**	**0.077**	**	**0.094**	**	**0.011**		**0.083**	**
Composition Effect attributable to																
Education	-0.025	**	-0.014	**	-0.013	**	-0.018	**	-0.036	**	-0.036	**	-0.022	**	-0.021	**
Rest	-0.061	**	-0.051	**	-0.039	**	-0.059	**	-0.061	**	-0.070	*	-0.061	**	-0.050	**
Error									-0.001		-0.011		-0.083	**	0.030	
Total explained by characteristics	**-0.086**	**	**-0.065**	**	**-0.052**	**	**-0.077**	**	**-0.098**	**	**-0.117**	**	**-0.165**	**	**-0.041**	+
Wage structure effects attributable to																
Education	-0.199	**	-0.151		-0.046		-0.078		-0.144	*	-0.291	**	-0.115	+	-0.042	
Rest	0.518	**	0.757	**	-0.025		0.116		0.474	*	1.349	**	0.048		-0.040	
Constant	-0.155		-0.448		0.134		0.122		-0.153		-0.851	*	0.241		0.207	
Error									-0.002		0.003		0.001		-0.001	
Total wage structure	**0.164**	**	**0.159**	**	**0.063**	**	**0.160**	**	**0.175**	**	**0.211**	**	**0.176**	**	**0.124**	**
Oriental	**A. Without reweighting**								**B. With reweighting**							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.093**	**	**0.103**	**	**0.082**	**	**0.074**	**	**0.093**	**	**0.103**	**	**0.082**	**	**0.074**	**
Composition Effect attributable to																
Education	0.019	**	0.010	*	0.010	**	0.014	**	0.017	**	0.012	*	0.014	**	0.020	**
Rest	0.035	**	0.038	**	0.032	**	0.036	**	0.035	**	0.022	+	0.027	*	0.054	**
Error									-0.001		-0.013		0.002		-0.014	
Total explained by characteristics	**0.054**	**	**0.049**	**	**0.042**	**	**0.050**	**	**0.052**	**	**0.021**		**0.043**	*	**0.059**	**
Wage structure effects attributable to																
Education	0.048		0.007		-0.085	+	-0.020		-0.010		-0.064		-0.100	+	-0.141	*
Rest	0.285		0.474	*	0.290	+	0.204		0.235		0.487	+	0.277		0.311	
Constant	-0.294		-0.426	+	-0.165		-0.160		-0.186		-0.344		-0.140		-0.156	
Error									0.002		0.003		0.003		0.001	
Total wage structure	**0.039**	*	**0.054**	*	**0.040**	*	**0.024**		**0.041**	*	**0.082**	**	**0.039**	*	**0.015**	

Notes: + p<0.1, * p<0.05, ** p<0.01.

Table 4.9 continue

Central	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.105**	**	**0.105**	**	**0.105**	**	**0.119**	**	**0.105**	**	**0.105**	**	**0.105**	**	**0.119**	**
Composition Effect attributable to																
Education	0.014	*	0.008	*	0.008	*	0.010	*	0.023	*	0.014	*	0.015	*	0.018	*
Rest	-0.008		0.008		-0.004		-0.008		-0.029	*	-0.021	+	-0.025	*	-0.026	*
Error									-0.001		-0.005		-0.006		0.002	
Total explained by characteristics	**0.006**		**0.015**		**0.004**		**0.002**		**-0.007**		**-0.012**		**-0.017**		**-0.007**	
Wage structure effects attributable to																
Education	-0.121	*	-0.088		-0.096	+	-0.097	+	-0.144	*	-0.102		-0.119	*	-0.120	*
Rest	0.044		0.236		0.169		-0.059		-0.009		0.127		0.083		-0.073	
Constant	0.175		-0.059		0.029		0.272		0.266		0.087		0.154		0.316	
Error									-0.001		0.005		0.002		0.002	
Total wage structure	**0.098**	**	**0.090**	**	**0.101**	**	**0.116**	**	**0.112**	**	**0.117**	**	**0.122**	**	**0.126**	**

Pacific	A. Without reweighting								B. With reweighting							
			Quantiles								Quantiles					
	OLS		25		50		75		OLS		25		50		75	
Overall wage gap	**0.433**	**	**0.535**	**	**0.433**	**	**0.333**	**	**0.433**	**	**0.535**	**	**0.433**	**	**0.333**	**
Composition Effect attributable to																
Education	0.009		0.005		0.005		0.006		-0.003		-0.002		-0.003		-0.002	
Rest	-0.022		-0.012		-0.017		-0.025	+	-0.032		-0.043		-0.042		-0.031	
Error									0.013		0.001		0.049		0.019	
Total explained by characteristics	**-0.014**		**-0.008**		**-0.012**		**-0.019**		**-0.022**		**-0.044**		**0.005**		**-0.014**	
Wage structure effects attributable to																
Education	-0.142	*	-0.092		-0.297	**	-0.194	*	-0.181	*	-0.133		-0.317	**	-0.208	*
Rest	0.285		0.503		0.602		-0.179		0.241		0.455		0.393		-0.264	
Constant	0.305		0.131		0.141		0.725		0.393		0.256		0.351		0.820	+
Error									0.002		0.001		0.001		0.000	
Total wage structure	**0.447**	**	**0.543**	**	**0.445**	**	**0.353**	**	**0.455**	**	**0.579**	**	**0.427**	**	**0.348**	**

Notes: + p<0.1, * p<0.05, ** p<0.01.

Figure 4.1. Regional hourly wage kernel density estimates - Thirteen largest metropolitan areas of Colombia

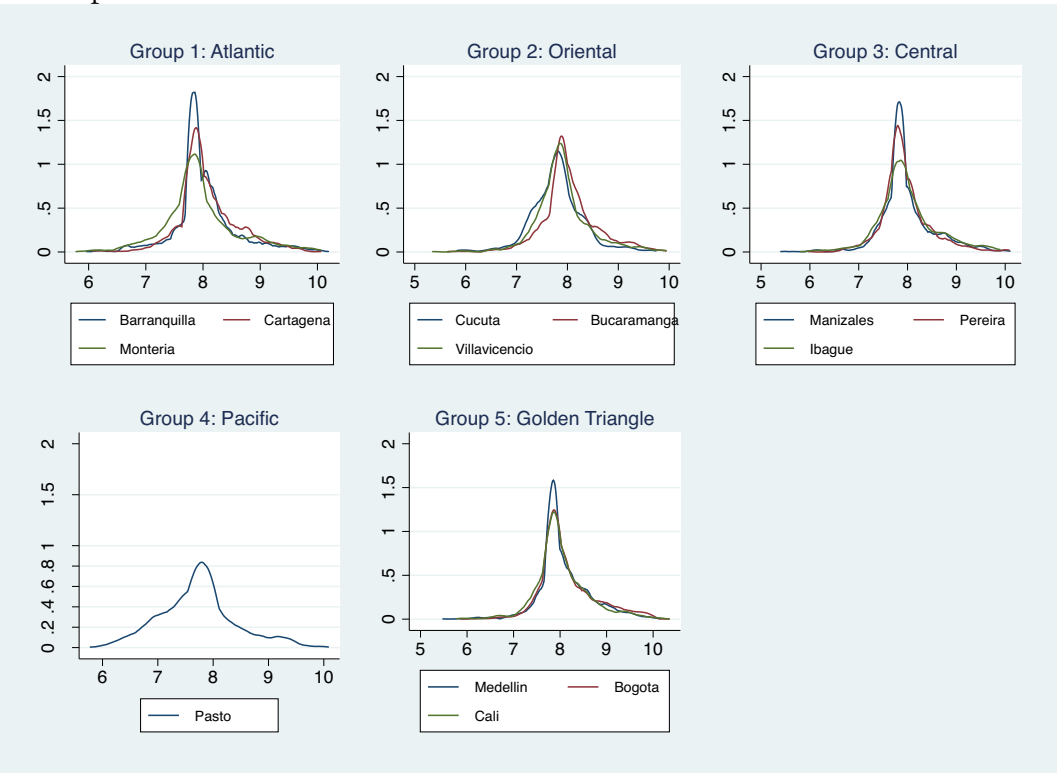


Figure 4.2. Regional hourly wage kernel density estimates - Five regions of Colombia

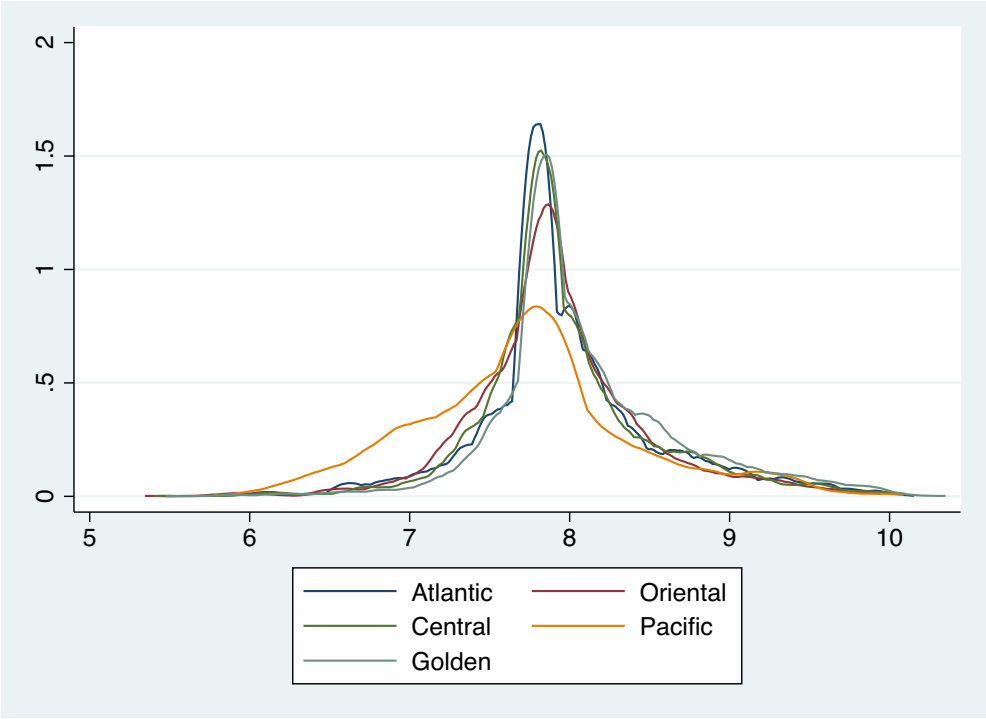
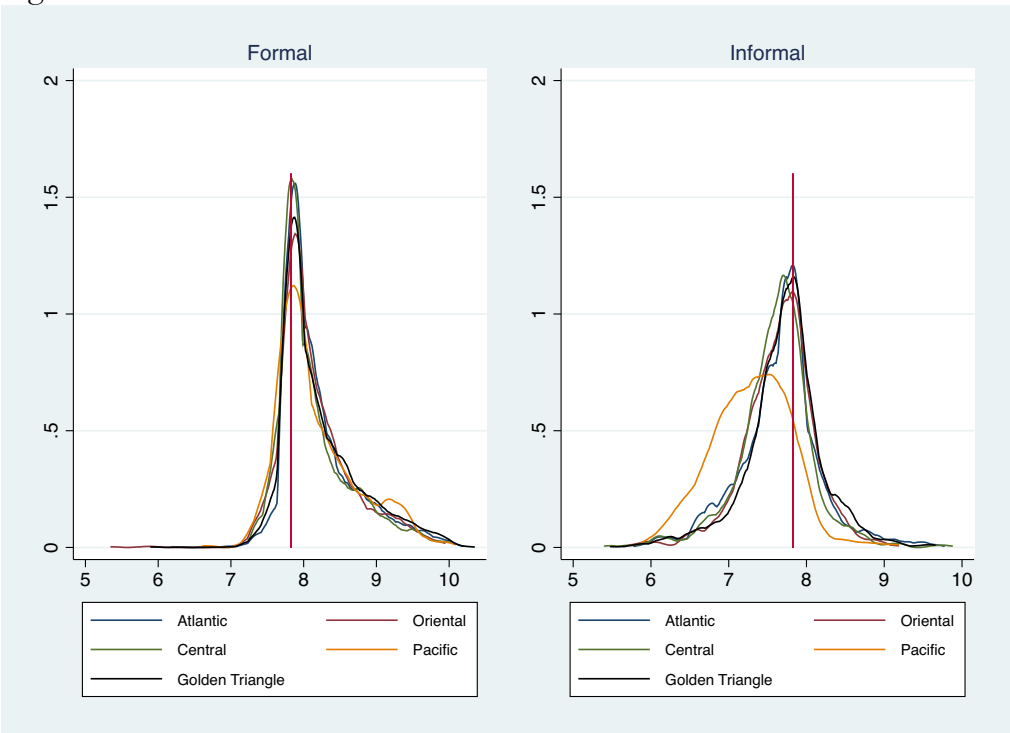


Figure 4.3. Formal and Informal hourly wage kernel density estimates - Five regions of Colombia



Chapter 5. General Conclusions

5.1 Main results and contributions

This PhD thesis represents new contributions to the empirical research of the functioning of the Colombian labor market, with a particular emphasis on education and its relation with informality.

In chapter 2 we analyzed if labor market segmentation, into formal and informal sectors, affects the allocation of educational skills for executing jobs. Particularly the case of overeducation is analyzed, the situation in which the education acquired by the worker exceeds the one required to perform his work. We estimated whether the probability that a worker is overeducated is determined by the sector of employment. Our results show that formal workers are less likely to be overeducated than their informal counterparts. We also found that workers are not randomly allocated into formal or informal jobs; and that unobservable characteristics that influence the probability of being assigned to a particular sector may also affect the probability of being overeducated, particularly for male workers. The results once this potential endogeneity of sector choice was taken into account, confirm that formal workers are less likely to be overeducated.

Motivated by the result obtained in chapter 2, chapter 3 explores the role of educational mismatches in explaining the formal-informal wage gap in Colombia. We found, as in previous studies, that formal workers have a higher return to their education, around double, compared with their informal counterparts. However, as a mayor contribution we demonstrate that, by including measures of educational mismatch important information to the analysis of the formal–informal wage gap is obtained. In particular, we showed that the returns to required education in the informal sector are not only lower, but the pay penalty that informal overeducated workers face in terms of wages are considerable higher than the one faced by their formal counterparts. The quantile regressions show that returns to required and surplus education are increasing along the wage distribution for formal workers. In contrast these returns are flat for informal workers. These results point that there is a second penalty associated with educational mismatches that puts informal workers at a greater disadvantage compare to formal workers. By using a decomposition analysis we revealed that while, under education is a minor

element for understanding the formal-informal gap in the returns to schooling, overeducation plays a more important role.

These results from the decomposition may also give some insight about which of the two most relevant theoretical frameworks may be more suitable for explaining why some characteristics, such as education, are better paid in the formal sector than in the informal sector in the Colombian case. Since under education is not what is driving the gap in returns to education between formal and informal workers, this result suggests that the lower returns to schooling for informal workers are probably not associated to sorting; workers with low levels of human capital are also those more likely to work in the informal sector. Rather, since overeducation is more relevant for explaining the gap, this result is more in line with the segmentation hypothesis, where some workers that do not have access to formal jobs are forced to accept informal jobs with lower education requirements and hence lower returns to their education.

According to the results found in chapters 2 and 3 it seems that, in addition to the benefits associated with receiving social security and earning higher wages, being a formal worker also ensures a better use of acquired skills in the workplace. To the best of our knowledge, no study has presented evidence of this to date.

Another explanation for our results is that educational mismatches, and the pay penalty associated with them, may be caused by lack of information on the part of job seekers. The lack of information entails that job searchers do not find or don't know how to search the job that is better suited for their education, while rigidities are factors that preclude them from getting the most appropriated job, even if it exists. The lack of information may be the result of the absence of appropriate job network links. 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, it is probable that education mismatching occurs for those individual who don't have access to these networks.

Chapter 4 is more self-contained. We analyze the role of education and informality for explaining regional wage differentials. The hypothesis of this study is that apart from the difference in the endowments of human capital across regions, regional heterogeneity in the incidence of informality may be another important source of regional wage inequality. Results showed the existence of differences not only in average regional wages but also across the

wage distribution. The decomposition analysis showed that for most of the regions and for most of the quantiles differences in the wage structures are the dominant factor explaining regional wage gaps. Meaning that workers with similar characteristics received different wages depending on the region in which they are located. At the middle and especially at higher quantiles of the wage distribution educations plays an important roll and a large part of wage differentials is due to difference in the returns to education. Informality and specifically its incidence only affect regional wage gaps at lower quantiles.

5.2 Policy recommendations

If labor market segmentation is what is contributing to the existence of overeducation in a developing country, then policies engaged with reducing informality could also have other positive effects apart from those commonly known, better quality jobs. Reducing informality may reduce the situation where a highly schooled worker takes a job with low-skill requirements and consequently a low pay. This evidence should be taken into consideration when assessing the issue of informality in the labor market of developing countries since it is likely to affect the allocation of skilled and unskilled workers in formal and informal jobs, and the incentives to accumulate education.

This study has also shown the importance of examining regional wage gaps separately for formal and informal workers; given that the results differ from those found when they are grouped together. Wage gaps are reduced considerably once formal workers are compared between regions, particularly for those regions with a high incidence of informality. Suggesting that formalization of employment, aside from the well-known implications of higher wages and social security coverage, may also help reducing disparities across regions. Moreover, if regional labor markets are segmented and formal and informal jobs are characterized by different mechanisms of functioning and adjustment, the proposed policy may not be unique for each of these two segments. Additionally policies that points towards the reduction of informality will help to lower regional wage gaps at the lower part of the wage distribution particularly for those regions with sizable informality. Likewise, reducing informality will help to reduce inter and intra regional wage inequality.

5.3 Limitation and future lines of research

This thesis has made important contributions to our understanding of the functioning of the labor market of a developing country. However, there are certain limitations that this thesis faced that are worth recognizing and which might be better overcome in future research.

First, it is worth mentioning the standard criticism on the suitability of instruments used for addressing endogeneity and sample selection in chapters 2 and 3. We use the presence of children in the household and the average number of years of schooling of other household members as instruments for sectorial assignment. These instruments were chosen based on data availability and because they had been used in other studies about informality (Günther and Launov, 2012). However there could be some reasons to suspect that these instruments might affect wages and the probability of overeducation. For instance, having a child might have an indirect effect on the offered wage through the reservation wage. Another criticism is that family conditions, such as the presence of children, since they may push the worker to accept an informal job, for the same reason it could happen that the worker may be forced to accept a not well-matched job. We intended to rule out this last fear with the robustness check conducted in section 2.5. Nevertheless, finding a credible instrument for sectorial choice has been proved to be a challenging task, as we didn't find any different instruments used in the extensive literature about informality. We think that the degree to which this limitation reduces the quality of our findings is a matter of debate. Moreover, we believe that our results are conclusive in terms of the correlations reported and might be a starting point to understand the importance of the effect of labor market segmentation on the probability of being overeducated.

Another common concern in the literature of overeducation is that low-ability workers might be incorrectly classified as overeducated because, all else equal, they do possess lower skill levels than high-ability workers. One way to deal with this problem is using panel data, which allows controlling for unobservable individual effects. Since ability does not vary within individuals, in principle, these estimates do not suffer from the same problem as the cross sectional ones. As mentioned by Leuven and Oosterbeek (2011) "fixed effects estimates of the returns to over/underschooling are identified from persons who have changed educational level, job level or both. In both cases it needs

to be the case that relevant unobservables are time-invariant." However if this unobservables are correlated with wages, fixed effects estimates will be biased as well.

Finally our results of chapters 2 and 3 may be related to the rise in commodity prices that is pushing Latin America toward another boom period led by its raw-materials exports. The increase in raw materials exports had boosted the exchange rates of these economies, making it harder for manufacturers to export. This extended boom in commodity prices has had a substantial reallocation of resources (including labor) from non-commodity tradable sectors (e.g. manufacture) to non-tradable sectors (e.g. services). Given that service sector is relatively less intensive in skilled labor and its activities are usually associated to informal jobs, this reallocation could have reduced the returns to education and can be responsible for the educational mismatch present in developing economies. We think that these two last hypotheses should be studied further in future research to get a better understanding about informality and educational mismatches in developing countries.

Regarding chapter 4, as in past studies of this nature, it remains to be explained why the difference in the returns to education across regions is persistent. We hypothesize that such a difference in returns is related to economies of scale and agglomeration economies; however further research is clearly required on this matter for a better understanding of regional wage differentials.

Bibliography

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

Alba-Ramírez, A. (1993). Mismatch in the Spanish labor market: Overeducation? *The Journal of Human Resources*, *28*, 259-278.

Albrecht, J., van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, *16*, 383–396.

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

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

Amuedo-Dorantes, C. (2004). Determinants and poverty implications of informal sector work in Chile. *Economic Development and Cultural Change, 52*, 347–368.

Andersson, F., Burgess, S., & Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, *61*, 112-128.

Arango, L. E., Herrera, P., & Posada C. E. (2008). El salario mínimo: Aspectos generales sobre los casos de Colombia y otros países. *Ensayos sobre Política Económica*, *26*, 204–263.

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

Azzoni, C. R. & Servo, L. M. (2002). Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science, 81*, 157-75.

Badaoui, E., Strobl, E., & Walsh, F. (2008). Is there an informal employment wage penalty? Evidence from South Africa. *Economic Development and Cultural Change*, *56*, 683-710.

Battu, H., Belfield, C., & Sloane, P. (2000). How well can we measure graduate overeducation and its effects? *National Institute Economic Review*, *171*, 82–93.

Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review* 21, 221-229.

Becker, Gary S. (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.

Blackaby, D. & Murphy, P. (1995). Earnings, unemployment and Britain's North–South divide: real or imaginary? *Oxford Bulletin of Economics and Statistics*, *57*, 487–512.

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

Bonet, J., & Meisel, A. (2007). Polarización del ingreso per cápita departamental en Colombia, 1975-2000. *Ensayos Sobre Política Económica*, *25*, 12-43.

Bonet J., & Meisel A. (2008). Regional Economic Disparities in Colombia, *Investigaciones Regionales*, *14*, 61–80.

Bosh, M., & Maloney, W. F. (2010). Comparative analysis of labor market dynamics using Markov processes: An application to informality. *Labour Economics*, *17*, 621–631.

Botelho, F. & Ponczek, V. (2011). Segmentation in the Brazilian labor market. *Economic Development and Cultural Change*, *59*, 437-463.

Büchel, F., & van Ham, M. (2003). Overeducation, regional labor markets, and spatial flexibility. *Journal of Urban Economics*, *53*, 482–493.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: A quantile regression approach. *Journal of Applied Econometrics*, *13*, 1-30.

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

Charlot, O., & Decreuse, B. (2005). Self-Selection in education with matching frictions. *Labour Economics*, *12*, 251-267.

Chevalier, A. (2003). Measuring over-education. *Economica*, *70*, 509-531.

Chiswick, B. R., & Miller, P. W. (2008). Why is the payoff to schooling smaller for immigrants? *Labour Economics*, *15*, 1317–1340.

Cohen-Zada, D., & Elder, T. (2009). Historical religious concentrations and the effects of catholic schooling. *Journal of Urban Economics*, *66*, 65-74.

Combes P., Duranton G. & Gobillon L. (2008). Spatial wage disparities: sorting matters! *Journal of Urban Economics*, *63*, 723–742.

Cruces, G., Garcia-Domenech, C. and Gasparini, L. (2011). Inequality in Education Evidence for Latin America. Mimeo, CEDLAS-UNLP and UNU-WIDER.

De la Rica, S., Dolado, J. J., & Llorens, V. (2008). Ceilings or floors? Gender wage gaps by education in Spain. *Journal of Population Economics*, *21*, 751–776.

Devillanova, C. (2013). Over-Education and spatial flexibility: New evidence from Italian survey data. *Papers in Regional Science*, *92*, 445–464.

Dinardo J., Fortin N.M. & Lemieux T. (1996). Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. *Econometrica*, *64*, 1001-1044.

Doeringer, P., & Piore, M. (1971). *Internal labor markets and manpower analysis*. Lexington, Mass.: Heath.

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. Graduate Labour Market. *Economics of Education Review*, *19*, 179-198.

Dominguez Moreno, J. A. (2009). Sobreeducación en el mercado laboral urbano de Colombia para el año 2006. *Revista Sociedad y Economía, 16*, 139-158.

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

Duranton G., & Monastiriotis V. (2002). Mind the gaps: the evolution of regional inequalities in the UK, 1982–1997. *Journal of Regional Science*, *42*, 219–256.

Duryea, S., Galiani, S., Nopo, H., & Piras, C. C. (2007). The educational gender gap in Latin America and the Caribbean. IDB Working Paper no. 502, Inter-American Development Bank, Washington, DC.

Evans, W. N., & Schwab, R. M. (1995). Finishing high school and starting college: Do catholic schools make a difference. *Quarterly Journal of Economics*, *110*, 947-974.

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

Fields, G. (1990). Labour Market Modelling and the Urban Informal Sector: Theory and Evidence, in: D., Thurnham, Salome, B. and Schwarz, A. (Ed.), *The Informal Sector Revisited*. Paris, OECD.

Fields, G. (2005). A Guide to Multisector Labor Market Models. Social Protection Discussion Paper Series, No 0505, World Bank, Washington, D.C., April.

Firpo S., Fortin N.M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*, 953- 973

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

Fortin N.M., Lemieux, T., & Firpo S. (2011). Decomposition Methods in Economics, in O. Ashenfelter and D. Card, eds., In *Handbook of Economics*, Amsterdam: North-Holland, Vol. IV.A: 1-102

Garcia, G. A. (2014). Labor informality: Choice or sign of segmentation? A quantile regression approach at the regional level for Colombia. MPRA Paper No. 55224. University Library of Munich, Germany.

Galvis, L. (2012). Informalidad laboral en las áreas urbanas de Colombia. *Coyuntura Económica*, *42*, 15-51.

García I., & Molina J. (2002). Inter-regional wage differentials in Spain. *Applied Economic Letters*, *9*, 209–215.

Garcia, J., Hernández, P. J., & López-Nicolás, A. (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics, 26*, 149–167.

Gasparini, L., Cruces, G. & Tornarolli, L. (2011). Recent Trends in Income Inequality in Latin America. *Economia*, *11*, 147-201 .

Gindling, T.H. (1991). Labor market segmentation and the determination of wages in the public, private-formal, and informal sectors in San José, Costa Rica. *Economic Development and Cultural Change*, *39*, 585.

Glaeser E., Kallal H., Scheinkman J. & Shleifer A. (1992). Growth in cities. *Journal of Political Economy*, *100*, 1126–1152.

Goldberg, P. K., & Pavcnik, N. (2005). Trade, wages, and the political economy of trade protection: Evidence from the Colombian trade reforms. *Journal of International Economics*, *66*, 75-105.

Gong, X., & Van Soest, A. (2002). Wage differentials and mobility in the urban labor market: A panel data analysis for Mexico. *Labour Economics*, *9*, 513-529.

Green, F., McIntosh, S., & Vignoles, A. (2002). The utilization of education and skills: Evidence from Britain. *The Manchester School*, *70*, 792-811.

Greenwood M. J., Hunt, G. L., Rickman D. S., & Treyz G. I. (1991). Migration, regional equilibrium, and the estimation of compensating differentials. *The American Economic Review*, *81*, 1382-1390 .

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

Günther, I., & Launov A. (2012). Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics, 97,* 88–98.

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

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

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

Hill, M. A. (1983). Female labor force participation in developing and developed countries-consideration of the informal sector. *The Review of Economics and Statistics*, *65*, 459-468.

International Labour Organization. (2011). 2011 Labour Overview: Latin America and the Caribbean. Geneva: International Labour Office.

Joumard, I., & Londoño Vélez, J. (2013). Income Inequality and Poverty in Colombia - Part 1. The Role of the Labour Market. OECD Economics Department Working Papers, No. 1036, OECD Publishing.

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

Kim, Y.-J. (2011). Catholic schools or school quality? The effects of catholic schools on labor market outcomes. *Economics of Education Review*, *30*, 546-558.

Koenker R., & Basset G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R. (2005). Quantile Regression. Cambridge Books, Cambridge University Press, number 9780521845731, November.

Kugler, A., & Kugler, M. (2009). Labor market effects of payroll taxes in a in developing countries: Evidence from Colombia. *Economic Development and Cultural Change, 57,* 335–358.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and Mismatch in the Labor Market. In *Handbook of the Economics of Education,* vol 4, ed. Eric A. Hanushek,

López-Bazo, E., & Motellón, E. (2012). Human Capital and Regional Wage Gaps. *Regional Studies*, *46*, 1347-1365.

López-Calva, L. F., & Lustig, N. (eds). (2010). Declining Inequality in Latin America: A decade of progress?. Washington, DC, The Brookings Institution.

Machado J., & Mata J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*, 445-465

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

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *World Bank Economic Review, 13,* 275–302.

Maloney, W. F. (2004). Informality revisited. *World Development*, *32*, 1159-1178.

Maloney, W. F., & Nu´ñez, J. (2004). Measuring the impact of minimum wages: Evidence from Latin America. In *Law and Employment: Lessons from Latin America and the Caribbean*, ed. James Heckman & Carmen Pagés, 109–130. Chicago: The University of Chicago Press.

Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. (1997). Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change*, *45*, 367-392.

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

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

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

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, *20*, 387-418.

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

Melly B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics* , *12*, 577-590

Mendes de Oliveira, M., Santos, M. C., & Kiker, B. F. (2000). The Role of Human Capital and Technological Change in Overeducation. *Economics of Education Review*, *19*, 199-206.

Mincer, J. (1974). Schooling, experience, and earnings. NBER, New York.
Mondragón-Veléz, C., Peña, X., & Wills, D. (2010). Labor market rigidities and informality in Colombia. *Economia*, *11,* 65-101.

Mora, J. J. (2005). Sobre educación en Cali (Colombia). ¿Desequilibrio temporal o permanente?: Algunas ideas, 2000-2003. Documentos Laborales y Ocupacionales, 2, SENA.

Motellón E., López-Bazo E. & Attar M. (2011). Regional heterogeneity in wage distributions: evidence from Spain. *Journal of Regional Science*, *51*, 558–584.

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

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

Organisation for Economic Co-operation and Development. (2011). OECD Employment Outlook 2011. OECD Publishing.

Ortiz, C., Uribe, J., & Badillo, E. (2008). Segmentación inter e intrarregional en el mercado laboral urbano de Colombia, 2001-2006. *Ensayos sobre Política Económica, 27*, 194-231.

Pereira, J. & Galego, A. (2013). Inter-regional wage differentials in Portugal: An analysis across the wage distribution. *Regional Studies*, DOI:10.1080/00343404.2012.750424

Pereira, J., & Galego, A. (2013). Decomposition of Regional Wage Differences Along the Wage Distribution in Portugal: the Importance of Covariates. CEFAGE-UE Working Paper no. 2013-16, University of Évora.

Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., & Saavedra-Chanduvi, J. (2007). Informality: Exit and exclusion. Washington, DC: World Bank.

Pisani, M. J., & Pagán J. A. (2004). Sectoral selection and informality: A Nicaraguan case study. *Review of Development Economics*, *8*, 541–556.

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

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

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

Quiñones-Domínguez, M., & Rodríguez-Sinisterra, J. (2011). Rendimiento de la educación en las regiones colombianas: un análisis usando la Descomposición Oaxaca-Blinder. *Sociedad y Economía*, *20*, 37-68.

Ramos, R., & Sanroma, E. (2012). Overeducation and local labour markets in Spain. *Tijdschrift voor economische en sociale geografie, 104*, 278–291.

Ren, W., & Miller, P. W. (2012). Gender Differentials in the Payoff to Schooling in Rural China. *The Journal of Development Studies*, *48*, 133–150.

Romero, J. (2008). Diferencias sociales y regionales en el ingreso laboral de las principales ciudades colombianas, 2001-2004. *Revista de Economía del Rosario*, *1*, 165–201.

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

Rumberger, R. (1987). The impact of surplus schooling on productivity and earnings. *Journal of Human Resources*, *22*, 24–50.

Sicherman, N. (1991). Overeducation in the labor market. *Journal of Labor Economics*, *9*, 101–122.

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

Tannuri-Pianto, M. E., Pianto, D. M., & Arias, O. (2004). Informal employment in Bolivia: A Lost proposition? World Bank, mimeo.

Tansel, A. (2000). Formal versus informal sector choice of wage earners and their wages in Turkey. In T. Bulutay (Ed.), Informal Sector (I) (125-146). Ankara: State Institute of Statistics, Printing Division.

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

Tokman, V. (1982). Unequal development and the absorption of labour: Latin America 1950–1980. CEPAL Review, 17, 121–33.

Tsang, M. C., Rumberger, R. W., & Levin, H. M. (1991). The impact of surplus schooling on worker productivity. *Industrial Relations: A Journal of Economy and Society*, *30*, 209-228.

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

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