

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Muriel Hernández, Beatriz

Working Paper An analysis of firm characteristics as earnings determinants: The urban Bolivia case

Development Research Working Paper Series, No. 04/2016

Provided in Cooperation with: Institute for Advanced Development Studies (INESAD), La Paz

Suggested Citation: Muriel Hernández, Beatriz (2016) : An analysis of firm characteristics as earnings determinants: The urban Bolivia case, Development Research Working Paper Series, No. 04/2016, Institute for Advanced Development Studies (INESAD), La Paz

This Version is available at: https://hdl.handle.net/10419/177356

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

Institute for Advanced Development Studies



4/2016

An Analysis of Firm Characteristics as Earnings Determinants: The Urban Bolivia Case

By:

Beatriz Muriel Hernández Institute for Advanced Development Studies

Development Research Working Paper Series 04/2016

December 2016

The views expressed in the Development Research Working Paper Series are those of the authors and do not necessarily reflect those of the Institute for Advanced Development Studies. Copyrights belong to the authors. Papers may be downloaded for personal use only.

An Analysis of Firm Characteristics as Earnings Determinants: The Urban Bolivia Case

by

Beatriz Muriel Hernández^{*} Institute for Advanced Development Studies 6115 Hector Ormachea Av., Obrajes, Phone: 491-2-2424065 La Paz -Bolivia <u>bmuriel@inesad.edu.bo</u>

Abstract

This article analyzes the importance of firm characteristics to explain earnings in urban Bolivia. Initially I propose a new simple theoretical model of segmented labor market where, in equilibrium, individual and firm variables jointly determine earnings at the worker level. The key for achieving this equilibrium is that workers have both specific preferences and heterogonous skills provided by years of schooling, which are in turn associated to certain firms. Given the household surveys information, I estimate two alternative earnings functions from this model, one for unsalaried workers, for which there is detailed firm data and one for salaried workers, in which sector, size and formality are used as firm proxies. I find not only that firm characteristics are fundamental determinants of earnings but that regressions that include only individual characteristics present highly overestimated coefficients.

Keywords: earnings functions, labor market segmentation, firm characteristics, Bolivia

JEL Classification: C26, J20, J24, J31

^{*} I would like to thank Albert Berry, Gustavo Canavire, Christopher Taber, Miguel Fernandez, Lykke Andersen and the INESAD team for the useful observations made to the draft version of this paper. Financial support from IDRC (International Development Research Centre) is also gratefully acknowledged. The usual disclaimer applies.

I. Introduction

Several empirical studies have been developed to analyze the factors that explain urban earnings in Bolivia. These investigations employ variables related to the Mincer earnings function, such as years of schooling and experience and they establish controls for gender and ethnic wage gaps (Pérez de Rada 1997, Fields *et al.* 1998, Rivero and Jiménez 1999, Andersen 2001, Andersen and Muriel 2002, Spatz and Steiner 2002, Ramírez 2003, Mercado *et al.* 2003, Escalante 2004 and Mercado and Aguilar 2006).

Nevertheless, Abowd *et al.* (1999) and Abowd and Kramarz (1999) highlight the importance of also considering the heterogeneous components of firm characteristics in the empirical analyses of wages. They demonstrate that the exclusion of these factors can lead to biases in the estimated parameters due to the omission of variables and/or aggregation problems.

The inclusion of firm characteristics in empirical estimations can be justified theoretically by relaxing an implicit hypothesis of the Mincer earnings function, which is that the labor market is perfect. In this regard, I propose a simple theoretical model of a segmented labor market where, in equilibrium, individual and firm variables jointly determine the earnings at the worker level. Furthermore, workers of two different segmented labor markets have different wages even if they have equal years of schooling. These results arise because individuals have heterogeneous preferences on both acquiring specific skills provided by education and working in a given segmented market. This model permits to avoid *ad hoc* approaches of earnings regressions where firm characteristics are included at the worker level, which have been broadly used in the literature.

From this model, and given the information available in Bolivia, I develop two alternative earnings regressions. The first regression is for unsalaried workers for which there is information on production costs in the household surveys, in addition to the usual individual variables. The second model is for salaried workers, where their firms' characteristics are approximated by the economic sector, firm size and formality. I highlight the relevance of these variables for explaining wages in the urban Bolivia case, thus supporting the alternative hypothesis of imperfections in the labor market.

The empirical analysis confronts three econometric problems: multicollinearity, endogeneity and sample selection bias. The multicollinearity arises for unsalaried workers because firm characteristics are derived from a *translog* production function, which has various variables related to the same inputs. I correct for this problem by using Bierrens' (2007) procedure and including only one representative variable of each input.

The endogeneity problem arises because workers' ability is an omitted variable in earnings regressions, which is correlated to years of schooling. Consequently, I use two instrumental variables to correct for the problem: A *dummy for computer usage* and a

dummy for semi-analphabets. Taken together, the two variables are strong predictors of years of schooling.

Last, the sample selection bias arises because of the separation of unsalaried workers and salaried workers from the labor force, which implies the use of non-randomly selected samples. I use the Wooldridge (2002) approach to correct for this bias while simultaneously correcting for the endogeneity problem, which combines Heckman's two-step procedure and instrumental variables.

This study represents a contribution to the empirical literature on earnings determinants in urban Bolivia for the following reasons. Firm characteristics are exposed as fundamental factors for explaining earnings for unsalaried workers. The coefficient of determination (R^2) doubles when these variables are included in the regression. In addition, the estimated coefficients of individual variables present as highly biased (overestimated) when firm characteristics are excluded, confirming the findings of Abowd *et al.* (1999) and Abowd and Kramarz (1999). For instance, the regression including firm characteristics shows that unsalaried workers have a rate of return of 3.4% per year of schooling, which represents only half of the corresponding estimated rates of return when firm variables are excluded.

For salaried workers, the inclusion of firm characteristics improves the coefficient of determination by more than 30%, showing that the economic sector, size of the firm and formality condition are important determinants of earnings. Similarly to that of unsalaried workers, the regression that only considers individual factors presents overestimated coefficients.

Finally, I estimate the Shorrocks (1982) and Fields (2002) inequality decomposition. The results show that firm characteristics jointly explain 27.6% of the earnings dispersion for unsalaried workers and 10.5% of the dispersion for salaried workers. In the former case, firm characteristics present higher explicative power than the individual variables together.

This article is organized as follows. Section II presents relevant stylized facts on earnings and their relationship with education and firm characteristics for urban Bolivia. Section III describes the theoretical model and the alternative earnings regressions derived, while Section IV presents the data, discusses the econometrical problems and shows the empirical estimations. Finally, Section V presents the most important conclusions.

II. Stylized Facts

The most important factor of incidence on earnings in urban Bolivia, as noted in the empirical literature, has been education. Years of schooling has an estimated return of approximately 8% to 9% per year and usually explains 50% or more of the explained

earnings inequality (Pérez de Rada 1997, Fields *et al.* 1998, Rivero and Jiménez 1999, Andersen 2001, Andersen and Muriel 2002, Spatz and Steiner 2002, Ramírez 2003, Mercado *et al.* 2003, Escalante 2004 and Mercado and Aguilar 2006).

Indeed, Figure 1 shows that average earnings by quintile have a positive relation with the corresponding average years of schooling. However, there is a high dispersion between these variables. For instance, an individual with a labor income corresponding to the third quintile may have between 6 and 15 years of education, while an employee with 8 years of schooling – completed primary school – may belong to any earnings quintile.





Other individual characteristics, such as experience, gender, ethnicity and educational quality, have been included in earnings regressions to explain the dispersion between earnings and years of schooling (Rivero and Jiménez 1999, Andersen and Muriel 2002, Mercado *et al.* 2003, Ramírez 2003 and Muriel 2005). Nevertheless, there are no studies that evaluate firm variables as determinants of earnings.¹ In this regard, I show three relevant firm characteristics at the individual level: economic sector, firm size and formality. Formality is evaluated through both the worker pension system affiliation and firm accounting records, which are highly correlated with the legal condition of firms in Bolivia.

Figure 2 shows the average earnings and the average years of schooling by economic sector. These variables still have a positive correlation, with a concentration of sectors in

Source: Prepared by author based on Bolivian household survey 2009.

¹ In some cases, however, economic sector dummies were considered.

both intensive qualified labor sectors² (over 12 years of education) and intensive unqualified sectors³ (with up to 12 years of schooling). However, some sectors present different earnings with similar levels of education, whereas others present similar earnings with different levels of education. For instance, workers in the mining-oil sector enjoy earnings similar to those of workers in the electricity, natural gas and water supply sector as well as those in the banking sector, but with fewer years of education, 9 years for the mining-oil sector compared to 13.4 and 15.3 years, respectively for the latter sectors. Furthermore, while construction and public administration have higher earnings compared with social and health services, social and health services shows more years of education. Finally, extraterritorial organizations and business services present a high wage gap even though they have similar levels of education.





Source: Prepared by author based on Bolivian household survey 2009. Note: The bubble size corresponds to the number of observations.

The high average earnings in construction and mining-oil are reflected in their higher production values during the last several years. For instance, in 2003-2004 both activities had average earnings lower than the education sector (see Figure A.1 in Annex). In contrast, commerce and manufacturing have been characterized by low labor productivities over time (see, e.g., Muriel and Jemio, 2010). Farming has low earnings because this household survey gather information during a sowing period (November-

² It includes electricity, natural gas and water supply, public administration, business services, social and health services, banking, education, and extraterritorial organizations.

³ It includes farming, hotels and restaurants, domestic service, construction, mining-oil, commerce, manufacturing, other social services, and transport and communication.

December), a period that requires a considerable amount of work but few returns (Andersen and Valencia, 2010).⁴

Table 1 shows the average earnings and the average years of education by firm size and formality, considering two groups according to Figure 2. The first group, called *qualified labor sectors*, includes sectors intensive in the use of workers with up to 12 years of schooling and the second group, *unskilled labor sectors*, includes the sectors comprised of the remaining workers.

In each group, average earnings increase with education, but the relation is less clear between groups. For instance, workers of medium and large firms belonging to the *unqualified labor sectors* have higher average earnings than those of micro-firms belonging to the *qualified labor sectors*, although the latter have more years of education. The formality category in the *unqualified labor sectors* has higher average wages than does informality in the *qualified labor sectors*, although education is lower in the first case.⁵

| | Unqualified | labor Sectors | Qualified labor Sectors | | |
|---------------------------|------------------------|---------------------|-------------------------|--------------------|--|
| | Ln(earnings) | Years of Schooling | Ln(earnings) | Years of Schooling | |
| Firm size | | | | | |
| Micro firms | 6.958 | 8.880 | 7.125 | 14.156 | |
| | (0.970) | (4.364) | (0.840) | (4.065) | |
| Small firms | 7.387 | 9.992 | 7.546 | 14.602 | |
| | (0.787) | (4.161) | (0.863) | (3.677) | |
| Medium-sized firms | 7.517 | 11.480 | 7.382 | 14.171 | |
| | (0.732) | (4.088) | (0.722) | (3.591) | |
| Large firms | 7.657 | 11.908 | 7.580 | 14.729 | |
| | (0.705) | (4.182) | (0.648) | (3.423) | |
| Formality (by both worker | pension system affilia | tion and firm accou | nting records) | | |
| Informal | 7.090 | 9.180 | 7.388 | 14.277 | |
| | (0.932) | (4.384) | (0.750) | (3.728) | |
| Formal | 7.749 | 13.118 | 7.822 | 15.837 | |
| | (0.638) | (3.720) | (0.644) | (2.444) | |

Table 1: Average of Ln(earnings) per Month and Average of Years of Schooling by Firm Size and Formality, 2009

Source: Prepared by author based on Bolivian household survey 2009. Note: In brackets are the standard errors.

In sum, the information above shows differences (or similarities) between earnings by economic sector, firm size and formality that are not explained by years of schooling. Abowd *et al.* (1999) and Abowd and Kramarz (1999) highlight the importance of these types of heterogeneous firm characteristics in the empirical analyses of earnings by demonstrating that the exclusion of these factors in earnings regressions can lead to high

⁴ Only the Household Survey 2003-2004 covered all the months of a year (November 2003-November 2004) and in this case the farming have higher average earnings than domestic services (see Figure A.1 in Annex). ⁵ All comparisons are supported by the Wald test.

biases in the estimated parameters, which are the result of omissions of variables and/or aggregation problems (the use of different linear combinations of variables).

The first important restriction to an adequate study of the factors that explain earnings rests on the need for information at the worker level of both individual and firm characteristics. These data have been available in a few cases, usually for developed countries or for specific studies (e.g. Verner 1999, Jones 2001, Soderbom *et al.* 2005, Card and de la Rica 2005, Martins 2008 and Aydemir and Skuterud 2008, Van Biesebroeck 2011). In other cases, empirical literature has been creative in matching individual and firm variables at the worker level. Some labor surveys with firm identifiers were matched with the corresponding establishment surveys. Others have data on the type of business and location, which allowed assigning an industry-location cell for every worker and matching it with the corresponding firm's information (e.g. Troske 1995, Abowd *et al.* 1999, Hellerstein *et al.* 1999, Moretti 2004, Hellerstein and Neumark 2004, Lallemand *et al.* 2005, Ilmakunnas and Maliranta 2005 and Breau and Rigby 2006).

Administrative records of workers have also been matched with the corresponding firm's information, which, in turn, is derived from either establishment surveys or tax statements (e.g., Haltiwanger *et al.* 2000, Menezes-Filho *et al.* 2006 and Woodcock 2008). Finally, some available firm variables, such as sector and size, have been included in earnings regressions (e.g. Brown and Medoff 1989, Oi and Idson 1999, Bayard and Troske 1999, Belfield and Weiz 2004, Temesgen 2005, Lallemand *et al.* 2005 and Muravyev 2009).

The second important shortcoming are the *ad hoc* specifications of earnings functions when firm variables are included, which raises doubts about how firm and individual characteristics are related to determine earnings. The exceptions are the models developed at the plant/firm level where two equations are estimated empirically: Mincer earnings functions where wages are averaged at the plant/firm level and productivity/production equations that include average employment characteristics (see Hellerstein et al. 1999, Crépon *et al.* 2003, Dostie 2006 and van Ours and Stoeldraijer 2011). Other models also are approximated empirically by including some specific variables in Mincer earnings functions at the individual level, as size, union bargaining, etc. (see, e.g., Lallemand *et al.* 2005, Gürtzgen 2009, El-Attar and López–Bazo 2008 and Muravyev 2009).

With regard to the employer-employee data availability I note that the Bolivian household surveys have ample information on labor variables at the individual level and include some firm characteristics, as discussed above. In addition, all production costs are available for unsalaried workers, which represent a considerable portion of the employed population, that is, 40% for 2009. This group consists of persons who are self-employed, managers or partners and members of production cooperatives who do not receive any salary, but have positive labor incomes. Establishment surveys, in contrast, are scarce for

the manufacturing sector and almost nonexistent for the rest of the industries, which limits any employer-employee matching.

With regard to the *ad hoc* specifications I propose a simple theoretical model that, in equilibrium, includes firm characteristics as additional earnings determinants at the worker level. From this model, I develop two alternative earnings regressions, one for unsalaried and one for salaried workers, according to the information available from the Bolivian household surveys.

III. The model

According to the supply-side of the labor market, earnings have been determined by education, which is derived from the human capital theory (Becker 1993 and Mincer 1974). This relationship assumes implicitly that the labor market is perfect, that is, two individuals with the same education have the same wage as well as the same marginal productivity value. Differences in education are then the source for explaining unequal wages and productivities. This theory has been refined supposing that marginal returns and costs (or tastes) of education are heterogeneous among workers, modeling abilities and considering other relevant individual characteristics (see Card 1999 and Heckman *et al.* 2003).

Nevertheless, when the labor market is not perfect wages can be different for workers with the same education or equal for workers with different years of schooling, even when controlling for other individual characteristics. Then, the demand-side of the labor market becomes relevant for determining earnings at the worker level. I support this premise by proposing a simple theoretical model that permits to determine a new earnings function at the worker level where both individual and firm characteristics are included simultaneously.

I suppose that the labor market is segmented in N groups of firms, where firms from group n (n=1, 2, ..., N) are identical. Workers have specific skills provided by education as well as particular preferences related with these groups. This specific skills assumption follows the models of workers with heterogeneous skills, which determine different occupational choices by worker in earnings regressions (see Roy 1951, Heckman and Sedlacek 1985, Keane and Wolpin 1997 and Sullivan 2007). In addition, workers with educational skills related to a group increase the productivity of firms belonging to this group. For simplicity, I assume initially that education is the only individual characteristics. As I show below, under these assumptions two different groups of firms n and n can pay different wages for the same level of years of schooling, but two firms k and k within a given group n pay the same wage for the same level of years of schooling.

Supply-side of the labor market

I follow Card (1995) to derive the Mincer earnings function through the utility maximization problem. Furthermore, I assume that there are *I* individuals from which a group of them I_n ($I_n = I_1, I_2, ..., I_N$, $0 < I_n < I$, $\sum_n I_n = I$, $I_n \neq I_n$.) prefer to acquire years of schooling related to specific skills needed in group *n* as well as to work in any firm belonging to this group. Then, an individual $i \in I_n$ choses the years of schooling that maximizes her utility function ($U(\cdot)$), which is described as:

(1)
$$U_{i \in I_n}(s(t_n), w_i(s)) = E[\ln w_i(s)\tau_i / s = s(t_n)] - h(s(t_n))$$

where $U(\cdot)$ has the usual properties, w(s) is the wage associated to *s* years of schooling, $\tau_i = \tau_{in} = 1$ if individual *i* works in group *n* and $0 < \tau_i = \tau_{in'} < 1$ if *i* works in group *n'*, $h(s(\iota_n))$ is a continuous twice differentiable convex function, and $s(\iota_n)$ are the years of schooling related to the specific skills demanded by group *n*, ι_n . The parameter τ_i makes explicit the disutility that an individual $i \in I_n$ has of working in a group of firms different from *n*, which may even be heterogeneous to any $n' \neq n$ as well as to any groups of individuals I_n . For simplicity, I assume that individual *i* always choses education with specific skills associated with her preferences, defined as: $h(s(\iota_n)) = \infty \forall n' \neq n$ for $i \in I_n$.

The first order condition of (1) is given by:

(2)
$$E\left[\frac{w'_i(s)}{w_i(s)}/s=s(\iota_n)\right]=h'(s(\iota_n))$$

which is a first-order differential equation in *s*. The Mincer earnings function is found considering that $h(s(t_n)) = rs(t_n)$ and resolving the differential equation:

(2')
$$E[w_i(s)/s = s(\iota_n)] = E[w_i(\bar{s})/\bar{s} = \bar{s}(\iota_n)]e^{r(s-\bar{s})}$$

where *r*>0 is the discount rate, $E[w_i(s)/s = s(\iota_n)] = \sum_n \rho_{in} w_{in}(s)$, $\sum_n \rho_{in} = 1$ and ρ_{in} is the probability that individual *i* can work in any firm belonging to group *n*.

Demand-side of the labor market

A representative firm k from group n pays an average wage $w_n(\bar{s})$ corresponding to the average of years of schooling of workers that belong to I_n ($\bar{s}(\iota_n)$), which equals the average labor marginal productivity value. This relationship is derived considering a

generic technology of production $F(A(\bar{s}(\iota_n)), \mathbf{x}_k)$, where $A(\cdot)$ shows that workers with specific skills provided by education $\bar{s}(\iota_n)$ can increase the productivity as a type of Hicks neutral technology and \mathbf{x}_k is the vector row of inputs. The function $A(\cdot)$ is a simple way of controlling for differences in average education of employment ($\in I_n$) between firms in group n, which leads to different productivities.

The optimal employment for the representative firm k is found through the resolution of the usual problem of profits maximization

(3)
$$w_n(\bar{s})|_{A=A(\bar{s}(\iota_n))} = p_n \frac{\partial F(A(\bar{s}(\iota_n)), \mathbf{x}_k)}{\partial x_{lk}} = \frac{p_n y_k}{x_{lk}} \frac{\partial \ln y_k}{\partial \ln x_{lk}} = \frac{p_n y_k}{x_{lk}} \varepsilon_{ln}$$

where y_k (= $F(\cdot)$) is the production of firm k, ε_{ln} is the employment-product elasticity and p_n is the price. Note that if the infinitesimal increase in employment is relevant to change the average educations of the specific skills in the firm from \bar{s} to $\bar{\bar{s}}$ ($\bar{s} \neq \bar{\bar{s}}$), (3) should be redefined as

$$w_n(\cdot) = w_n(\bar{s})\big|_{A=A(\bar{s}(\iota_n))} + p_n \frac{\partial F(\cdot)}{\partial A(\bar{s}(\iota_n))} \frac{\partial A(\bar{s}(\iota_n))}{\partial \bar{s}(\iota_n)} \frac{\partial \bar{s}(\iota_n)}{\partial x_{lk}} = w_n(\bar{s})\big|_{A=A(\bar{s}(\iota_n))}.$$

Equilibrium

The equilibrium in the labor market depends on the matching of the specific skills between the demand and supply of employment. A perfect match implies that: $\sum_{k\in n} x_{lk} = I_n / \varphi \ \forall n \in N$, where φ redefines the unities of employment to be comparable between the demand and the supply sides. In this case all individuals that chose specific skills related to group n in their corresponding education are hired by firms from this group ($\rho_{in} = 1 \ \forall i \in I_n$ in (2')) and all these firms only contract workers that have education related with their specific skills requirements. Furthermore, preferences and wages are such that $\ln w_{in}(s(t_n)) \ge \ln w_{in'}(s(t_n)) \tau_{in'} \ \forall i \in I_n, \forall n' \neq n \in N$. In equilibrium there is no workers mobility between different groups of firms, individuals work in the groups of firms according to their corresponding preferences and firms contract individuals that can increase their corresponding productivities.

I find the equilibrium wage at the worker level by assuming that firm k pays to each worker a wage considering not only the average labor marginal productivity value but also the worker's education. This equilibrium is established as a proposition.

Proposition: The wage of worker i ($i \in I_n$) with s (= $s(\iota_n)$) years of schooling who works in firm k from group n is determined as:

(4)
$$w_{in}(s) = w_n(\bar{s})g_{in}(s-\bar{s}) = w_n(\bar{s})e^{r(s-\bar{s})}$$

where the function $g_{in}(s-\bar{s})$ values the years of schooling gap between worker i and employment (on average) as a result of the marginal productivity gap between them. The equality $g_{in}(s-\bar{s}) = e^{r(s-\bar{s})}$ is derived from the equilibrium in the labor market.

Proof: Define $g_{in}(s-\bar{s})$ as weights that allow adding workers in "productivity equivalent units"⁶ derived from different years of schooling between them. Employment can then be redefined as $x_{lk} = \sum_{i} x_{ilk} = \sum_{i} g_{in}(\cdot)x_{ilk}$ (condition 1), where x_{ilk} is the labor supply given by worker *i*.

To illustrate this point, assume that two workers, *i* and *i*', offer one unit of labor. If education is higher for the former (*s*) than for the latter (*s*'), then in "productivity equivalent units," worker *i* represents more labor units than worker *i*' and $g_{in}(s-\bar{s}) > g_{i'n}(s'-\bar{s})$ (condition 2). For instance, one unit of labor of *i* can represent double that one unit of *i*', which implies $g_{in}(\cdot) = 2g_{i'n}(\cdot)$.

The wage for worker *i* is derived from the maximization problem considering that her education changes $A(\cdot)$ and that $\breve{s}(\iota_n)$ is the previous workers' average years of schooling

(5)
$$w_{in}(s) = \left[p_n \frac{\partial F(\cdot)}{\partial x_{lk}} + p_n \frac{\partial F(\cdot)}{\partial A(\cdot)} \frac{\partial A(\bar{s}(\iota_n))}{\partial \bar{s}(\iota_n)} \frac{\partial \bar{s}(\iota_n)}{\partial x_{lk}} \right] \frac{\partial x_{lk}}{\partial x_{ilk}} = \frac{p_n y_k}{x_{lk}} \varepsilon_{ln} g_{in}(\cdot) = w_n(\bar{s}) \Big|_{A=A(\bar{s}(\iota_n))} g_{in}(\cdot)$$

where $\frac{\partial x_{lk}}{\partial x_{ilk}} = g_{in}$ by condition 1.

The labor supply and demand in group *n* imply that equations (2') and (5) are equal in equilibrium once the individual *i* is contracted in firm *k* ($\rho_{in} = 1$). This implies that $g_{in}(s - \bar{s}) = e^{r(s-\bar{s})}$, which meets conditions 1 and 2.

In addition, for two firms, *k* and *k*' in group *n*, with average wages of $w_n(\bar{s})$ and $w_n(\bar{s})$, respectively, the following must be met in the labor market equilibrium: $w_n(\bar{s})g_{in}(s-\bar{s}) = w_n(\bar{s})g_{i'n}(s-\bar{s})$. This ensures that both firms pay the same wages to workers with equal years of schooling and implies that $A(s(\iota_n)) = e^{rs(\iota_n)}$ for $s = \bar{s}, \bar{s}$.⁷

Last, for firms belonging to two different groups, n and n', $w_n(s) \neq w_{n'}(s)$ in equilibrium, which follows directly from assumptions that $I_n \neq I_{n'}$, $\ln w_{in}(s(t_n)) \ge \ln w_{in'}(s(t_n))\tau_{in'}$. This equilibrium means that the labor market is composed by N sub-labor markets where an individual $i \in I_n$, with $s(t_n)$ years of schooling, is not willing to offer her work to a group $n' \neq n$ because her utility level would be equal or even worse under this choice.

An imperfect match implies then that $\ln w_{in}(s(t_n)) < \ln w_{in'}(s(t_n))\tau_{in'} \quad \forall i \in I_n$ in some groups of firms \tilde{n} (< *N*). Then individuals with preferences related to groups \tilde{n} are willing to offer

⁶ This approximation follows Trefler (1993).

⁷ Of course under the assumption of identical firms education should be the same, $\bar{s} = \bar{s}$, but this is not a necessary condition for this equilibrium.

their work in other groups \breve{n} $(\breve{n} + \widetilde{n} \le N)$ because the disutility of working in these last groups is compensated by higher wages. In turn, wages from groups \breve{n} will decrease to adjust the excess of labor supply and wages from groups \widetilde{n} will increase (given the reduction in the labor supply). This process will continue until the equilibrium between the *N* sub-labor markets is restored. Note that multiple equilibriums can be obtained depending principally on all I_n and τ_i values.

The wage of an individual $i \in I_n$ that works in a firm k' from group n' is also represented by equation (4), with the difference that $p_n \cdot \frac{\partial F(\cdot)}{\partial A(\bar{s}(\iota_n, \cdot))} \frac{\partial A(\bar{s}(\iota_n, \cdot))}{\partial \bar{s}(\iota_n, \cdot)} \frac{\partial \bar{s}(\iota_n, \cdot)}{\partial x_{lk'}} = 0$ for an infinitesimal increase of employment of individual I.

The model's assumptions can be approximated to the reality. They say that workers have preferences for acquiring specific skills provided by years of schooling, which explain why a country have doctors, teachers, civil engineers, industrial engineers, etc. These specific skills are directly related with the technology of firms. For instance, a petroleum engineer will be more productive than a doctor in the oil sector, even though the years of schooling are the same between them. In addition, the hypothesis that $I_n \neq I_{n'}$ says that there is not an equal number of teachers, petroleum engineers, nurses, etc. and surely most of them will experience disutility working in firms that do not require their corresponding specific skills provided by their years of schooling.

Certainly there are some years of schooling that are not specific to any group of firms (such as secretaries, custodians, managers, etc.). For these cases the model can be relaxed supposing that there is a group of individuals $I_{\bar{n}}$ that have preferences for acquiring different kind of skill provided by education that are not directly related with any group n, but who have heterogeneous preferences of working in a specific group. These assumptions generate the imperfect match equilibriums described above.

Econometric specifications derived from the model

Given that $A(\bar{s}(\iota_n)) = e^{r\bar{s}(\iota_n)}$ in equilibrium, I redefine expression (4) and take logarithms to obtain

(4') $\ln w_{in}(s) = \ln \overline{w}_n(\overline{s}) + \ln e^{rs}$

where $\tilde{w}_n(\bar{s})(=w_n(\bar{s})/A(\bar{s}(\iota_n)))$ is determined by the firm characteristics according to (3). From equation (4') two alternative earnings regressions are constructed according to the available data for urban Bolivia.

The part of the wage related to observable individual factors, $\ln e^{rs}$ in (4'), is the wellknown Mincer earnings function. I relax the assumption that workers are different only by education and include other individual relevant differences, where *rs* represents now a function of these variables such that

(5)
$$\ln e^{rs(\cdot)} = \alpha_0 + \alpha_1 s_i + \alpha_2 e x_i + \alpha_3 e x_i^2 + \sum_m \xi_m d_{im}$$

where s_i is years of schooling, ex_i and ex_i^2 are experience and squared experience, respectively, d_{im} is the other m^{th} relevant attribute of individual *i* for determining wages (and productivity) and the alphas and xis are the coefficients to be estimated.

For unsalaried workers, the part of the wage related with observable firm factors are directly obtained from (3). Assuming that $F(\cdot)$ is a *translog* production function, I take logarithms to (3) and replace this new expression with (5) into (4'), which yields the following:

(6)
$$\ln w_{in}(s) = \alpha_0 + \gamma_1 \ln \varepsilon_{in} + \sum_j \gamma_j \ln x_{jk} + \frac{1}{2} \sum_j \sum_{j'} \gamma_{jj'} \ln x_{jk} \ln x_{j'k}$$
$$+ \alpha_1 s_i + \alpha_2 e x_i + \alpha_3 e x_i^2 + \sum_m \xi_m d_{im} + u_i$$

where sub-indices j, j' are attributed to inputs; u_i is the error term; $\gamma_{jj'} = \gamma_{j'j} \forall j$, j' and $j \neq j'$; $\sum_j \gamma_{jj'} = 0$; $\sum_{j'} \gamma_{j'j} = 0$; and $\sum_j \frac{\gamma_j}{\varpi} = 1 - \varpi$ represents the degree of homogeneity of the production function (Berndt and Christensen 1973 and Binswanger 1974).

The second regression corresponds to salaried workers where wages are determined from (4') as follows:

(7)
$$\ln w_{in}(s) = \widetilde{\alpha}_0 + \alpha \ln \widetilde{w}_n(\overline{s}) + \widetilde{\alpha}_1 s_{1i} + \widetilde{\alpha}_2 s_{2i} + \widetilde{\alpha}_3 e x_i + \widetilde{\alpha}_4 e x_i^2 + \sum_m \widetilde{\xi}_m d_{i,m} + \upsilon_i$$

where the alphas and xis are the coefficients to be estimated, υ_i is the error term with the usual properties and $\ln \omega_n(\bar{s}) \approx function(sector, size of firm, formality)$. As showed above, these variables segment the Bolivian urban labor market, which would be related with both heterogonous preferences and specific skills provided by education, experience, etc.

Equations (6) and (7) are simplified for further discussion as

(8)
$$\widetilde{w}_{in}(s) = \sum_{c} \beta_{c} \widetilde{x}_{ic} + v_{i}$$

where $\tilde{w}_{in}(s) = \ln w_{in}(s)$, β_c (*c*=0, 1, 2,.., *C*) is the estimated coefficient corresponding to the explanatory variable \tilde{x}_{ic} (*c*=0, 1, 2,.., *C*) and v_i is the error term.

Finally, I evaluate the main determinants of earnings inequality. For this, I use the decomposition analysis methodology of Shorrocks (1982) and Fields (2002), which consists of estimating the variance components of $\tilde{w}_{in}(s)$ in (8), normalized to 1 (or 100%), yielding the following expression (see also Cowell and Fiorio 2009):⁸

$$(8') \qquad 1 = \sum_{m} \theta_{m} + \theta_{v}$$

$$\theta_{m} = \beta_{m}^{2} \frac{VAR(\tilde{\boldsymbol{x}}_{m})}{VAR(\tilde{\boldsymbol{w}})} + \sum_{m' \neq m} \beta_{m} \beta_{m'} \frac{COV(\tilde{\boldsymbol{x}}_{m}, \tilde{\boldsymbol{x}'}_{m'})}{VAR(\tilde{\boldsymbol{w}})} + \beta_{m} \frac{COV(\tilde{\boldsymbol{x}}_{m}, \boldsymbol{v}')}{VAR(\tilde{\boldsymbol{w}})},$$

$$\theta_{v} = \frac{VAR(\boldsymbol{v})}{VAR(\tilde{\boldsymbol{w}})} + \sum_{m} \beta_{m} \frac{COV(\tilde{\boldsymbol{x}}_{m}, \boldsymbol{v}')}{VAR(\tilde{\boldsymbol{w}})} \text{ and } \tilde{\boldsymbol{x}}_{m}, \tilde{\boldsymbol{w}} \text{ ,and } \boldsymbol{v} \text{ are row vectors.}$$

IV. Empirical Analysis

IV.1. Econometric Issues

The estimation of regressions (6) and (7) presents three potential econometrics problems: multicollinearity (for unsalaried workers), endogeneity and sample selection bias. The multicollinearity problem arises because several explanatory variables derived from the *translog* production function are highly correlated, given the cross products between them, which tend to inflate the standard deviations (deflating the *t*-values) as well as the coefficient of determination R^2 . I correct the problem by using Bierrens' (2007) procedure, which consists of i) choosing the least relevant variable for the analysis that is also insignificant; ii) regressing this variable over the rest of the explanatory variables; iii) estimating the residuals (which are not correlated with the rest of the explanatory variables); iv) replacing the explanatory variable chosen in the first step by the residuals estimated in the main regression, which re-parameterizes the model; and v) if there are some variables that are still insignificant, repeating the procedure until all the *t*-values of the remaining variables are significant.

The endogeneity problem arises because – according to the human capital theory – worker's ability is an omitted variable (unobserved) that is correlated with the years of

⁸ This methodology has been criticized when the dependent variable is in logarithms, which may break up (possibly in a rare case) the transfer axiom: if a new distribution is obtained from another one by taking income from the lower-paid worker to the better-paid worker, measured inequality should increase. However, I use this method because as far as I know there is not, as yet, a better method for regression-based inequality decomposition (for a discussion of the methodologies, see Morduch and Sicular 2002, Cowell and Fiorio 2006, and Israeli 2007).

schooling, which accordingly, leads to a biased estimated coefficient for education.⁹ In this regard, instrumental variables are used, taking as instruments the institutional features of the school system or family background, such as the education of spouses (see Card (1999) and Card (2001) for surveys). Given the information from the Bolivian household surveys, I propose two instruments. The first instrument is a dummy equal to one for workers who used a computer at least once during the last twelve months. The second instrument is a dummy equal to one for workers who cannot read and write, or who learned to read and write through the literacy program "Yo Si Puedo", implemented in Bolivia during 2006 to 2008.¹⁰ Both variables capture diverse restrictions and limitations of access to schools, universities computer courses and lack of money, which are not properly related with abilities.¹¹ Both dummy variables are jointly strong predictors of years of schooling (see Table A.2 in Annex).

Empirically, an additional equation is added for years of schooling:

(9)
$$s_i = \delta_0 + \delta_1 Z_{1i} + \delta_1 Z_{2i} + \eta_i$$

where z_{1i} and z_{2i} are the *dummy for computer usage* and the *dummy for semianalphabet*, respectively, and η_i is the error term with the usual properties.

The last econometric problem, sample selection bias, arises because the separation of unsalaried workers and salaried workers from the working-age population implies the use of non-randomly selected samples for missing data problems (see Heckman, 1979). To correct for this bias, the literature usually uses Heckman's two-step procedure. However, because the equations analyzed include two biases at the same time – endogeneity and sample selection – I use the Wooldridge (2002) approach, which combines Heckman's two-step procedure and instrumental variables.¹² Initially, a probit model is estimated for unsalaried workers (or salaried ones) from the entire urban working-age population (10 or more years of age according to the Bolivian institute of statistics), as a function of all

⁹ However, where workers have heterogeneous costs and returns, other kinds and causalities of biases arise (see Card, 1999).

¹⁰ This program was created by the Latin American and Caribbean Pedagogical Institute, based in Havana, Cuba. It is an audiovisual method that lasts approximately two months.

¹¹ Computer usage had limitations because computer rooms/classes appeared in Bolivia at the end of the eighties in some universities and private schools, mostly located in the Bolivia main cities: La Paz, Santa Cruz and Cochabamba. Over time these rooms/classes were increasing, but only since 1997 the students' for teachers institutions had access to them and since the two thousand years some public schools (which cover around 90% of the student population) had computer rooms/classes. According to the household survey 2009, 75% and 40% of students in private and public schools, respectively, used a computer during the last twelve months. Alphabetization had limitations because of lack of schools principally in the rural areas, lack of money and culture attitudes restricting mainly women to education. For instance, in 1992 the analphabetism rate was 20%, 28% for women and 37% for rural areas (see Table A.1 in Annex).

¹² From a previous use of this methodology see Garcia *et al.* (2001), Das *et al.* (2003), Reza and Mussurov (2006), and Chen and Hamori (2009).

relevant characteristics, some of them not included in (8) and excluding the years of schooling:

(10)
$$P(q_i = 1 | \widetilde{\widetilde{\mathbf{x}}}_i) = \Phi(\boldsymbol{\zeta}' \widetilde{\widetilde{\mathbf{x}}}_i)$$

where q_i is equal to 1 if the individual in the working age population is an unsalaried (or salaried) worker and is between 18 and 65 years of age and zero otherwise and $\tilde{\tilde{x}}_i$ is a vector of the explanatory significant variables.

Next, the coefficients are estimated using all observations, thus permitting the calculation of the inverse Mills ratio $\hat{\lambda}_i = \lambda(\hat{\boldsymbol{\zeta}}' \widetilde{\boldsymbol{x}}_i)$ for the sub-sample of the working-age population studied. Expression (8) is then redefined including the inverse Mills ratio:

(8")
$$\widetilde{w}_{is,kn} = \sum_{c} \beta_{c} \widetilde{x}_{ic} + \beta_{\lambda} \hat{\lambda}_{i} + v_{i}$$

Finally, equations (9) and (8") are jointly estimated for both unsalaried and salaried workers.

IV.1. Data Description

The information required for empirically analyzing equations (9) and (8') was obtained from the 2009 Bolivian household survey. However, in a draft version of this article, I used the household surveys of 2003-2004 and 2007 with similar econometric conclusions (see Muriel 2011). Table 2 describes both individual and firm variables.

| Table 2: Individual and Firm Characteristics | , Bolivian Household Survey | 2009 |
|--|-----------------------------|------|
|--|-----------------------------|------|

| <i>Ln</i> (earnings) by month | This is calculated by the Bolivian institute of statistics (INE). In the case of unsalaried workers, it corresponds to the net profit by deducting all declared costs (payment of salaries, intermediate consumption, service expenses, tax payments and other contributions) from the gross income. In the case of salaried workers, it is equal to the basic wage minus taxes and contributions plus all the additional monetary and nonmonetary payments (overtime, bonuses, food, transport, etc.). Earnings are not converted into working hours because this variable had unreliable data in many observations (e.g., people working 16 hours or more per day during all the days of the month). |
|-------------------------------|--|
| Ln(hours_month) | The logarithm of hours of work by month, which is included for controlling monthly earnings dispersion due to differences on them. |
| Years of schooling | Worker's years of schooling. |
| Age | Worker's age. It is used as proxy of the usual experience variable, given that many workers become workers before finishing their education. |
| D_male | Dummy equal to one for men. |
| D_indigenous | Dummy for indigenous people, measured by the first language. |

| Table 2 continua | tion |
|------------------|------|
|------------------|------|

| Additional individua | al characteristics for the Heckman's two-step procedure |
|--|---|
| Children | The number of children alive. It is expected to be positively related with unsalaried workers because child care requires greater flexibility in working hours. Furthermore, this variable characterizes people with lower years of schooling as well as economic opportunities in the formal labor market. |
| D_household head | A dummy equal to one for the household head, which reinforces the probability of belonging to the employed population from the working-age population. |
| D_migrant | A dummy for migrant people, which is positively correlated with the employed population given that many migrants change regions when looking for labor opportunities. |
| D_student | A dummy for those who enrolled the year of the survey, expecting a negative relation with any work condition (salaried and unsalaried). |
| Specific variables cooperatives who do | for unsalaried workers (self-employed, managers or partners and members of production not receive any salary but have positive labor incomes) |
| Ln(product- employment elasticity) by sector | This is obtained noting that the elasticity is equal to the share of the labor costs over the total cost (see, e.g., Yasar <i>et al.</i> 2008). It is approximated by the ratio between monthly earnings plus other labor costs and monthly gross income and averaged for all workers from a given economic sector. |
| Ln(employment) | This is the employment in the unsalaried workers production unit. |
| Ln(capital) | This is a proxy variable for capital corresponding to the monthly payments for services such as rent, water, electricity, phone, security, etc. |
| Ln(intermediate consumption) | It sums up the monthly purchase of raw materials, tools, equipment and merchandise. |
| Dummies by sector | There are 13 sector dummies that have unsalaried workers: farming, mining-oil, commerce, electricity, natural gas and water supply, construction, manufacturing, hotels and restaurants, transport and communication, banking, business services, education, social and health services and other social services. These variables are included to control relevant sector characteristics, such as the market power coefficient described in equation (6). |
| Specific variable for | r salaried workers (workers that receive wages) |
| Dummies by sector and firm size | Dummies by economic sector and firm size. The 16 sectors include besides those previously described, public administration, extraterritorial organizations and domestic services. Firm size is divided into four categories, according to INE: micro (1 to 4 workers), small (5 to 14), medium (15 to 49) and large (50 or more workers). Sizes were accounted for all sectors except for electricity, natural gas and water supply, banking, public administration, education, social and health services, domestic service and extraterritorial organizations. Electricity, natural gas and water supply and banking were not divided by firm size because of too few observations, nor were the remaining sectors because they were subject to specific wage policies. In addition, small firms for mining-oil were excluded given the lack of observations. The division resulted in 42 groups for the entire sample. |
| D_pension affiliation | A dummy equal to one for people who declared to be affiliated with the pension system. It is a proxy of formal jobs found in formal firms, which is associated with both better productive conditions and wages. |
| D_accounting records | A dummy equal to one for people who declared that their firms have accounting records (totally or partially). It is a proxy for formal firms, which have accounting records for paying taxes as well as better management. |

The variables were constructed for workers ages 18 to 65 who declared a positive labor income (excluding two extreme values). Summary statistics for unsalaried and salaried workers are provided in Table 3.

| | Unsalaried | Workers | Salaried Workers | | |
|---|------------|-----------------------|------------------|-----------------------|--|
| Variable | Mean | Standard Deviation | Mean | Standard Deviation | |
| | | | | | |
| <i>Earnings</i> (in Bs. = 0.143 US\$) | 1865.749 | 2650.550 | 2071.438 | 1940.947 | |
| Ln(earnings) | 6.997 | 1.055 | 7.368 | 0.740 | |
| Years of schooling | 9.061 | 4.742 | 11.913 | 4.473 | |
| Proportion: dummy semi-analphabet workers | 0.070 | | 0.025 | | |
| Proportion: dummy for computer usage | 0.257 | | 0.568 | | |
| Experience | 41.179 | 11.598 | 34.991 | 11.400 | |
| Proportion of males | 0.511 | | 0.621 | | |
| Proportion of indigenous | 0.308 | | 0.141 | | |
| Hours of work per month | 212.358 | 105.387 | 199.689 | 77.653 | |
| Product-employment elasticity | 0.616 | 0.150 | | | |
| Employment (in persons) | 7.329 | 130.627 | | | |
| Capital (in Bs. = 0.143 US\$) | 114.757 | 352.279 | | | |
| Intermediate consumption (in Bs. = 0.143 USD) | 2588.460 | 12001.640 | | | |
| Proportion: dummy for pension affiliation | | | 0.362 | | |
| Proportion: dummy for accounting records | 1 | | 0.386 | | |
| Children | 0.996 | 1.695 | 0.565 | 1.262 | |
| Proportion: dummy for household head | 0.576 | | 0.491 | | |
| Proportion: dummy for student | 0.056 | | 0.161 | | |
| Proportion: dummy for migrant | 0.057 | | 0.076 | | |
| Number of observations | 142 | 1 | 2 ⁻ | 159 | |

Table 3: Summary Statistics, Bolivian Household Survey 2009

IV.2. Econometric Results

Table 4 shows the econometric results of earnings determinants for unsalaried workers. The first three regressions include only the observable individual characteristics: regression (1) is the OLS estimation, (2) corrects for the endogeneity bias and (3) corrects for endogeneity and sample selection biases at the same time. The coefficients estimated show the expected signs in all cases, being consistent with those previously found in the literature. That is, *years of schooling* has a positive effect on earnings and *age* also presents a positive impact with decreasing marginal returns. In addition, the coefficients of gender and ethnicity show that males earn more than females and that indigenous workers earn less than non-indigenous workers.

Next, the multicollinearity problem is evaluated for firm characteristics. The variables ln(capital), $ln(capital) \times ln(employment)$ and $ln(capital) \times ln(intermediate consumption)$ are discarded by using the Bierrens (2007) procedure. In addition, I consider only one representative variable of intermediate consumption as well as employment, which have the highest *t*-values. I follow this procedure because first and second order effect variables are highly correlated, which may overestimate both the relevance of firm characteristics and the coefficient of determination used for the inequality decomposition analysis.

| Explanatory variables | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------------------|------------|------------|--------------------|------------|------------|--------------------|
| | OLS | IV (2SLS) | IV plus Heckman | OLS | IV (2SLS) | IV plus Heckman |
| Years of schooling | 0.037 | 0.058 | 0.064 | 0.016 | 0.027 | 0.033 |
| C C | (0.006)*** | (0.010)*** | (0.011)*** | (0.006)*** | (0.010)*** | (0.010)*** |
| Age | 0.056 | 0.056 | 0.041 | 0.030 | 0.030 | 0.016 |
| | (0.015)*** | (0.015)*** | (0.016)** | (0.013)** | (0.013)** | (0.014) |
| Age ² | -0.0006 | -0.0006 | -0.0004 | -0.0003 | -0.0003 | -0.0002 |
| | (0.000)*** | (0.000)*** | (0.000)** | (0.000)** | (0.000)** | (0.000) |
| D_male | 0.411 | 0.371 | 0.353 | 0.409 | 0.392 | 0.378 |
| | (0.054)*** | (0.055)*** | (0.056)*** | (0.049)*** | (0.050)*** | (0.050)*** |
| D_indigenous | -0.299 | -0.220 | -0.261 | -0.256 | -0.219 | -0.257 |
| | (0.062)*** | (0.071)*** | (0.072)*** | (0.051)*** | (0.058)*** | (0.059)*** |
| Ln(hours of work) | 0.461 | 0.473 | 0.462 | 0.290 | 0.300 | 0.290 |
| | (2.986)*** | (2.696)*** | (-0.263)*** | (0.998)*** | (0.969)*** | (-0.249)*** |
| Ln(product-employment | | | | 0.998 | 0.969 | 0.952 |
| elasticity) by sector | | | | (0.128)*** | (0.128)*** | (0.128)*** |
| Ln(employment) | | | | 0.175 | 0.175 | 0.171 |
| | | | | (0.044)*** | (0.043)*** | (0.042)*** |
| [In(capital)] ² | | | | 0.004 | 0.003 | 0.003 |
| | | | | (0.002)*** | (0.002)** | (0.002)** |
| [In(intermediate | | | | 0.018 | 0.018 | 0.018 |
| consumption)] ² | | | | (0.001)*** | (0.001)*** | (0.001)*** |
| D_farming | | | | -0.941 | -0.917 | -0.914 |
| | | | | (0.124)*** | (0.125)*** | (0.124)*** |
| D_manufacturing | | | | -0.447 | -0.444 | -0.455 |
| | | | | (0.061)*** | (0.061)*** | (0.061)*** |
| D_banking | | | | 0.591 | 0.514 | 0.439 |
| | | | | (0.084)*** | (0.10)*** | (0.106)*** |
| Inverse Mills ratio | | | -0.263 | | | -0.249 |
| | | | (0.095)*** | | | (0.080)*** |
| Constant | 2.986 | 2.696 | 3.349 | 4.439 | 4.256 | 4.859 |
| | (0.339)*** | (0.362)*** | (0.431)*** | (0.306)*** | (0.331)*** | (0.385)*** |
| R^2 | 0.22 | 0.21 | 0.21 | 0.44 | 0.44 | 0.44 |
| Observations | 1421 | 1421 | 1421 | 1421 | 1421 | 1421 |

Table 4: Ln(earnings) Determinants for Unsalaried Workers, Bolivian Household Survey 2009

Notes: 1) In brackets are the standard errors calculated from the robust variance-covariance matrix; 2) *** means statistical significance at 1% and ** at 5%; 3) the *probit* model (for calculating the inverse Mills ratio) includes the following significant variables: *D_indigenous*, *D_household head*, *D_student*, *Children* and *D_migrant*.

Regressions (4) through (6) add the firm characteristics. Regression (4) shows the estimation using OLS and regressions (5) and (6) correct for the endogeneity bias and simultaneously for both biases, respectively. As above, all coefficients of individual characteristics present the expected signs. However, the variables *age* and *age*² are no longer significant when the Mills ratio is included. This result may be due to some Mills

ratio predictors, principally the dummies for student and household head, which are correlated with $age.^{13}$

In the last regression, I highlight the following results derived from the inclusion of firm characteristics. *Ln(employment)* shows a positive and relatively high coefficient, which can be associated with the size of the firm; an increase in 1% in employment is associated to an earnings increase of 0.17%. The *proxy* variable of capital has a positive effect on earnings; however, the elasticity is relatively low at 0.01 ($0.003 \times 2 \times ln(capital)$).¹⁴ This low elasticity may be due to a low use of this factor of production as well as the *proxy* used that can underestimate this coefficient. The earnings-intermediate consumption elasticity has a high value; an increase of 1% in these expenses is associated with an increase of 0.20% in labor income.¹⁵

Last, the wage gaps by economic sectors are significant and robust to alternative specifications in farming, manufacturing and banking. The dummy for farming is partially explained, as previously mentioned, by the few incomes perceived during the period of the survey. The manufacturing variable reflects, among other things, the low productivity. Finally, the dummy for banking may be controlling the market power of this sector.

The inclusion of firm characteristics in earnings regressions is exposed as fundamental for unsalaried workers as two main differences arise by comparing regressions (3) with (6). First, the coefficient of determination (R^2) doubles, from 0.21 to 0.44, when firm variables are included. Second, *years of schooling* is exposed as highly biased when firm characteristics are excluded, confirming the findings of Abowd *et al.* (1999) and Abowd and Kramarz (1999) regarding the variable omission problems.¹⁶ In particular, regression (3) predicts that an additional year of education would increase earnings by 6.6%, ($e^{0.064}$ -1)×100, while in (6), the estimated increase reaches only 3.4% (approximately half of the former).

The fact that the coefficients of individual characteristics are overestimated when firm characteristics are excluded can be explained by the relationship between the two types of variables (see Abowd *et al.* 1999). The exclusion of firm variables implies that the estimated coefficients of individual variables sum both effects, individual plus (employment weighted average) firm effects, when the covariance matrix between these two categories of variables is not zero. For instance, unsalaried workers with higher education usually belong to wealthier families and/or may have greater savings from

¹³ See Wooldridge (2002) for a discussion of this kind of problem.

¹⁴ *Ln(capital)* corresponds to the average sample equal to 1.572.

¹⁵ In this case, elasticity is equal to $0.034 \times 2 \times ln$ (*intermediate consumption*), where the variable corresponds to the average sample equal to 5.495.

¹⁶ The education coefficients of regressions (3) and (6) are statistically different according to the Wald Test (at 1% of significance).

previous jobs, which places them in an advantageous position to accumulate inputs, both in quantity and quality.

Regressions (7) through (12) in Table 5 present the econometric results for salaried workers. The last three regressions include firm characteristics, which have been narrowed down with dummies by firm's sector-size and formality. The dummies included are those that remained statistically significant at 10% under all the alternative combinations between them in the econometrical estimations. Regressions (8) and (11) were corrected by the endogeneity problem and (9) and (12) by both endogeneity and sample selection biases. All coefficients of individual characteristics are significant in most of the cases.

| Explanatory variables | (7) | (8) | (9) | (10) | (11) | (12) |
|---------------------------------|------------|------------|----------------------|---------------------|---------------------|----------------------|
| | OLS | IV (2SLS) | IV plus Heckman | OLS | IV (2SLS) | IV plus Heckman |
| Years of schooling | 0.051 | 0.069 | 0.075 | 0.035 | 0.048 | 0.054 |
| | (0.003)*** | (0.005)*** | (0.006)*** | (0.007)*** | (0.010)*** | (0.009)*** |
| Age | 0.063 | 0.057 | 0.047 | 0.045 | 0.043 | 0.034 |
| | (0.008)*** | (0.008)*** | (0.008)*** | (0.009)*** | (0.008)*** | (0.008)*** |
| [Age] ² | -0.0006 | -0.0005 | -0.0004 | -0.0004 | -0.0004 | -0.0003 |
| | (0.0)*** | (0.0)*** | (0.0)*** | (0.0)*** | (0.0)*** | (0.0)*** |
| D_male | 0.241 | 0.234 | 0.123 | 0.232 | 0.226 | 0.122 |
| | (0.029)*** | (0.029)*** | (0.037)*** | (0.041)*** | (0.041)*** | (0.042)*** |
| D_indigenous | -0.103 | -0.035 | 0.071 | -0.098 | -0.062 | 0.036 |
| | (0.036)*** | (0.039) | (0.045) | (0.040)** | (0.048) | (0.043) |
| Ln(hours of work) | 0.476 | 0.515 | 0.491 | 0.465 | 0.491 | 0.468 |
| | (0.037)*** | (0.037)*** | (0.037)*** | (0.044)*** | (0.045)*** | (0.046)*** |
| D_pension affiliation | | | | 0.313 (0.050)*** | 0.270 (0.045)*** | 0.254 (0.044)*** |
| D_accounting records | | | | 0.098 (0.021)*** | 0.085 (0.023)*** | 0.091 (0.023)*** |
| Inverse Mills ratio | | | -0.309 (0.061)*** | | | -0.285 (0.041)*** |
| Constant | 2.724 | 2.391 | 3.051 | 3.271 | 3.021 | 3.612 |
| | (0.243)*** | (0.254)*** | (0.281)*** | (0.258)*** | (0.269)*** | (0.302)*** |
| Dummies by sector and firm size | No | No | No | Yes | Yes | Yes |
| R ² | 0.29 | 0.28 | 0.29 | 0.38 | 0.37 | 0.38 |
| Observations | 2159 | 2159 | 2159 | 2159 | 2159 | 2159 |

Table 5: Ln(Earnings) Determinants for Salaried Workers, Bolivian Household Survey 2009

Notes: 1) In brackets are the standard errors calculated from the robust variance-covariance matrix and corrected by intra-groups correlation associated to sector and firm size division (42 clusters); 2) *** means statistical significance at 1% and ** at 5%; 3) the *probit* model (for calculating the inverse Mills ratio) includes the following significant variables: *D_male*, *D_indigenous*, *D_household head*, *D_student* and *D_migrant*; 4) The dummies by sector and firm size significant to alternative specifications are nine: large firms in mining, micro and small firms in manufacturing, small firms in construction, micro firms in commerce, hotels-restaurants and transport-communication, medium-sized business services and extraterritorial organizations.

The inclusion of the firm's variables improves the coefficient of determination by more than 30%. In the last regression, I highlight the relevance of formality for explaining earnings. The dummy for affiliation with the pension system has a positive and high impact, increasing the median earnings by approximately 31.4%, which can be related to better firm production conditions. Firms with accounting records also pay better wages, thus increasing the median earnings by 7.7%, which can be associated with better management.

Similar to that of the unsalaried workers, the regressions that only consider individual characteristics present statistically overestimated coefficients.¹⁷ In particular, an increase in one year of schooling is associated to an earnings increase of 7.8% in regression (9) and 5.6% in the last regression, which includes firm variables.

The overestimation of coefficients is explained by the relationship between individual and firm characteristics. For instance, *years of schooling* presents a positive correlation with the size of firms, showing that firms use more educated workers as they increase in size. In addition, the variable *D_pension affiliation* is positively related with *years of schooling* and *age*, which suggests that formal firms tend to hire workers with both more education and experience.

In addition, I analyze the robustness of the previous estimations by i) changing the estimation method used to limited-information maximum likelihood, ii) including nonlinearities in *years of schooling*,¹⁸ iii) including instrumental variables for a possible endogeneity between ln(employment) and ln(earnings) for unsalaried workers regressions and iv) considering all dummies by sector and firm size (41 dummies) for salaried workers regressions (see Tables A.3 in Annex).

Finally, Table 6 shows the earnings inequality decompositions from the estimated regressions that control for both endogeneity and sample selection biases. Regressions with individual characteristics only – (3) and (9) – present comparable results to previous empirical findings where education was the most relevant variable (Fields *et al.* 1998, Andersen 2001, and Spatz and Steiner 2002).¹⁹However, the inclusion of firm characteristics in earnings regressions – (6) and (12) – provides different conclusions. These variables jointly explain 27.6% of the earnings dispersion for unsalaried workers and 10.5% of the dispersion for salaried workers; thus, firm characteristics are key sources

¹⁷ The education coefficients of (9) and (12) are statistically different (at 1% of significance) according to the Wald Test.

¹⁸ Nonlinearities in education have been important in some econometrical findings for the urban areas of Bolivia (e.g., Moensted 2000, Muriel 2005, and Muriel 2011).

¹⁹ These studies excluded the residuals from the inequality decomposition analysis.

of inequality. Furthermore, in the case of unsalaried workers, firm characteristics have higher explicative power than do individual variables considered jointly.

| Explanatory variables | Unsalaried | Workers | Salaried Workers | | |
|-------------------------------|----------------|----------------|------------------|-----------------|--|
| | Regression (3) | Regression (6) | Regression (9) | Regression (12) | |
| Years of schooling | 5.92% | 3.38% | 9.16% | 7.88% | |
| Age plus [Age] ² | 0.64% | 0.28% | 6.92% | 5.20% | |
| D_male | 4.25% | 5.00% | 1.57% | 1.72% | |
| D_indigenous | 1.91% | 2.07% | -0.20% | -0.13% | |
| Ln(hours of work) | 7.86% | 5.42% | 7.59% | 8.85% | |
| Inverse Mills ratio | 0.23% | 0.23% | , 3.49% | 3.63% | |
| Firm characteristics together | | 27.64% | , | 10.47% | |
| Residuals | 79.19% | 55.97% | , 71.47% | 62.38% | |
| Total | 100.00% | 100.00% | 100.00% | 100.00% | |

Table 6: Shorrocks/Fields Earnings Inequality Decompositions, Bolivian Household Survey 2009

V. Conclusions

This article analyzes the relevance of firm characteristics as earnings determinants for the urban areas of Bolivia. Initially, I discuss some stylized facts showing that there are differences (or similarities) between earnings by economic sector, firm size and formality that are not explained by years of schooling, which is the most important explanatory variable found in the Bolivian empirical literature.

In this regard, I develop a new simple theoretical model that produces a new earnings function where both individual and firm characteristics are included. I assume that the labor market is segmented in groups of firms, where firms in each group require specific skills provided by years of schooling in order to increase their productivities. In turn, individuals have heterogeneous preferences on acquiring these specific skills as well as on working in a given group. In this model, the Mincer earnings function becomes valid only as a supply-side equation, which is matched to the demand equation at the worker level in order to achieve the equilibrium. The results show that two different groups of firms can pay different wages for the same level of years of schooling, which is consistent with the stylized facts for the urban Bolivia.

Given the information available in the Bolivian household surveys, I derive two alternative earnings regressions from the proposed model. The first model corresponds to unsalaried workers, which have data on production costs, thus allowing a detailed evaluation of earnings determinants. The second model is for salaried workers, where firm characteristics are approximated by economic sector, firm size and formality.

The empirical analysis shows that firm characteristics are fundamental factors for explaining earnings in the case of urban Bolivia. The coefficient of determination (R^2) improves significantly, doubling for unsalaried workers regressions and increasing by

more than 30% for salaried ones. Furthermore, individual variables are exposed as highly biased when firm characteristics are excluded, thus confirming the findings of Abowd *et al.* (1999) and Abowd and Kramarz (1999) regarding the variables omission problems. In particular, the estimations show the rate of return per year of schooling is 3.4% for unsalaried workers and 5.6% when firm variables are included.

Finally, the Shorrocks/Fields earnings inequality decomposition shows that firm characteristics are fundamental factors for explaining earnings dispersion. Furthermore, in the former case, firm characteristics have higher explicative power than individual variables considered jointly.

Annex



Figure A.1.: Ln(earnings) per Month by Sector and Years of Schooling, 2003-2004 (Employed population between 18 and 65 years of age)

| | 1976 | 1992 | 2001 | 2009 | | | |
|---|--------------|-------|-------|-------|--|--|--|
| Analphabetism rate (population of 15 years old or more) | | | | | | | |
| Total | 36.8 | 20.0 | 13.3 | 8.8 | | | |
| Men | 24.2 | 11.8 | 6.9 | 4.2 | | | |
| Women | 46.8 | 27.7 | 19.4 | 13.3 | | | |
| Urban areas | 15.6 | 8.9 | 6.3 | 4.7 | | | |
| Rural areas | 53.2 | 36.5 | 25.8 | 17.8 | | | |
| Population | | | | | | | |
| Total | 100.0 | 100.0 | 100.0 | 100.0 | | | |
| Urban areas | 41.2 | 57.5 | 62.4 | 66.2 | | | |
| Rural areas | 58.8 | 42.5 | 37.6 | 33.8 | | | |
| Dummy for semi-analphabet (equal to 1) in urban a | nd rural are | as | | | | | |
| 7-17 years old | | | | 1.2 | | | |
| 18-39 years old | | | | 6.2 | | | |
| 40-65 years old | | | | 21.9 | | | |
| More than 65 years old | | | | 47.6 | | | |

Table A.1: *Analphabetism* and Population Indicators, 1976-2009 (In percentages)

Source: Prepared by author based on the Bolivian institute of statistics data: 1976, 1992 and 2001 census of population and 2009 household survey.

Source: Prepared by author based on Bolivian household survey 2003-2004. Note: The bubble size corresponds to the number of observations.

| Table A.2: OLS Regressions for Years of S | Schooling, Bolivian Household Survey 2009 |
|---|---|
| | |

| Explanatory variables | Unsalarie | d Workers | Salaried Workers | | |
|--------------------------------|---------------------------|---------------------|-----------------------------------|---------------------|--|
| | (1) | (2) | (3) | (4) | |
| Dummy for semi-analphabet | -6.469 | -5.280 | -7.164 | -5.895 | |
| Dummy for computer usage | (0.323)*** | (0.338)*** 3.990 | (0.534)*** 5.110 (0.157)*** | (0.540)*** 3.197 | |
| Additional controls | (0.221) No | (0.241)**** Yes | (0.157)**** No | (0.244)*** Yes | |
| R ² Observations | 0.42 1421 | 0.51 1421 | 0.43 2159 | 0.56 2159 | |

Notes: 1) In brackets are the standard errors calculated from the robust variance-covariance matrix; 2) *** means statistical significance at 1%; 3) the additional controls included are those corresponding to regression (6) from Table 4 for unsalaried workers and regression (12) from Table 5 from salaried ones.

Table A.3: Robustness Checks, Bolivian Household Survey 2009

| Alternative estimations compared with | Regression (3) | Regression (6) | Regression (9) | Regression (12) |
|---|------------------|------------------|------------------|------------------|
| Limited-information maximum likelihood estimation method | | | | |
| Statistical differences between coefficients ⁽¹⁾ | No | No | No | No |
| R ² | 0.21 | 0.44 | 0.29 | 0.39 |
| Nonlinearities in years of schooling | | | | |
| Significance of the coefficient of years of | Insignificant at | Insignificant at | Insignificant at | Insignificant at |
| schooling with over 12 years (0 otherwise) | 10% | 10% | 1% | 10% |
| Instrumental variable for <i>In(employment</i>): A dummy for medium and large firms size (15 workers or more) | | | | |
| Statistical differences between coefficients ⁽¹⁾ | | No | | |
| R ² | | 0.44 | | |
| Inclusion of all firm characteristics: D_pension affiliation, D_accounting records, 41 dummies by sector and size | | | | |
| Statistical differences between coefficients ^{(1) (2)} | | | | No |
| R ² | | | | 0.39 |

Notes: ⁽¹⁾ statistical differences between coefficients are determined using the Wald test,⁽²⁾ statistical differences between dummies by sector and size were not considered nor the insignificant coefficient of $D_{indigenous}$.

References

Abowd, John M., and Francis Kramarz. 1997. Internal and External Labor Markets: An Analysis of Matched Longitudinal Employer-Employee Data. Working Paper no. 6109, National Bureau of Economic Research, Cambridge, MA.

Abowd, John M. and Francis Kramarz. 1999. The Analysis of Labor Markets using Matched Employer-Employee Data. In *Handbook of Labor Economics* (volume 3B), ed. Orley C. Ashenfelter and David Card, North-Holland: Elsevier.

Abowd, John M., Francis Kramarz and David N. Margolis. 1999. High Wage Workers and High Wages Firms. *Econometrica* 67, no. 2 (March): 251 – 333.

Abowd, J. M., F. Kramarz, D. and S. Woodcock. 2006. Econometric Analyses of Linked Employer-Employee Data. Cornell University, Centre de Recherche en Economie et Statistique (CREST), Simon Fraser University.

Andersen, Lykke E. 2001. Wage Differentials between Bolivian Cities. *Economía Política* II, no. 1: 48-62.

Andersen, Lykke E., and Beatriz Muriel H. 2002. Cantidad versus Calidad en Educación: Implicaciones para Pobreza, *Revista de Estudios Económicos y Sociales* 1: 8-41.

Aydemir, Abdurrahman, and Mikal Skuterud. 2008. The Immigrant Wage Differential within and across Establishments. *Industrial & Labor Relations Review* 61, no. 3: 333-352.

Bayard, Kimberly and Kenneth R. Troske. 1999. Examining the Employer-Size Wage Premium in the Manufacturing, Retail Trade and Services Industries using Employer-Employee Matched Data. *American Economic Review* 89, no. 2: 99-103.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* 3rd ed. Chicago: University of Chicago Press.

Belfield, R. Clive, and Xiangdong Weiz. 2004. Employer Size–wage Effects: Evidence from Matched employer–employee Survey Data in the UK. *Applied Economics* 36: 185–193.

Berndt, Ernst R., and Laurits R. Christensen. 1973. The Translog Function and the Substitution of Equipment, Structures and Labor in U.S. Manufacturing 1929–68. *Journal of Econometrics* 1 (March): 81–114.

Bierrens, Herman J. 2007. Multicollinearity. Department of Economics, Pennsylvania State University.

Binswanger, Hans P. 1974. The Measurement of Technical Change Biases with Many Factors of Production. *American Economic Review* 64, no. 6: 963-976.

Breau, Sébastien, and David L. Rigby. 2006. Is There Really an Export Wage Premium? A Case Study of Los Angeles Using Matched Employee – Employer Data. *International Regional Science Review* 29, no. 3: 297-310.

Brown, Charles, and James Medoff. 1989. The Employer Size-Wage Effect. Working Paper no. 3336, National Bureau of Economic Research, Cambridge, MA.

Cain, Glenn. 1976. The Challenge of Segmented Labor Market Theories to Orthodox Theory. *Journal of Economic Literature*, 14 (121):5-1257.

Card, David (1995), "Earnings, schooling and ability revisited", in: Solomon Polachek, ed., *Research in labor economics*, Vol. 14 (JAI Press, Greenwich, CT) pp, 23-48.

Card, David. 1999. The Causal Effect of Education on Earnings. In *Handbook of Labor Economics* (volume 3A), ed. Orley C. Ashenfelter and David Card. North-Holland: Elsevier.

Card, David. 2001. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 69, no. 5 (September): 1127-1160.

Card, David, and Sara de la Rica. 2005. The Effect of Firm-Level Contracts on the Structure of Wages: Evidence from Matched Employer-Employee Data. Working Paper no. 11829. National Bureau of Economic Research, Cambridge, MA.

Chen Guifu, and Shigeyuki Hamori. 2009. Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach. *China Economic Review* 20: 143–152.

Cowell, Frank A., and Carlo V. Fiorio. 2006. Rethinking Inequality Decomposition: Comment. London School of Economics, University of Milan and Econpublica.

Cowell, Frank A., and Carlo V. Fiorio. 2009. Inequality Decompositions: A Reconciliation. Working Paper no. 117, Society for the Study of Economic Inequality, London.

Crépon, Bruno, Nicolas Deniau and Sébastien Pérez-Duarte. 2003. Wages, Productivity and Worker Characteristics: A French Perspective". Working Paper CREST-INSEE, Paris.

Dostie Benoit (2006), "Wages, Productivity and Aging," IZA Discussion Paper no. 2496, Bonn, Germany.

Dickens, T. William, and Kevin Lang. 1992. Labor Market Segmentation Theory: Reconsidering the Evidence. Working Paper no. 4087. National Bureau of Economic Research, Cambridge, MA.

El-Attara Mayssun and Enrique López-Bazo. 2008. "Firm Size and the Entire Distribution of Wages: A Semiparametric Approach". *Applied Economics Letters* (15): 427–430.

Escalante, Scarlet. C. 2004. Los Retornos de la Inversión en Capital Humano en Bolivia, *Revista de Análisis Económico* 19: 1-26.

Fernández-Huerga, Eduardo. 2010. La Teoría de la Segmentación del Mercado de Trabajo: Enfoques, Situación Actual y Perspectivas de Futuro. *Investigación Económica* LXIX(273): 115-150.

Fields, Gary S. 2002. Accounting for Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States, *Articles & Chapters*, paper no. 265.

Fields, Gary S., Jesse B. Leary, Luis Felipe López-Calva, and Ernesto Pérez de Rada. 1998. Education's Crucial Role in Explaining Labor Income Inequality in Urban Bolivia, Development Discussion Paper no. 658, Harvard Institute for International Development, Cambridge, MA.

Gould, Eric. 2002. "Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States." *Journal of Labor Economics* 20(1): 105-47.

Haltiwanger, John C., Julia I. Lane, and James R. Spletzer. 2000. Wages, Productivity and the Dynamic Interaction of Businesses and Workers, Working Paper no. 7994, National Bureau of Economic Research, Cambridge, MA.

Hamermesh, Daniel S. 2008. Fun with Matched Firm – Employee Data: Progress and Road Maps. *Labour Economics* 15: 663–673.

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47, no. 1: 153-161.

Heckman, James and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self Selection in the Labor Market." *Journal of Political Economy* 93(6): 1077-1125.

Heckman, James and Guilherme Sedlacek. 1990. "Self-Selection and the Distribution of Hourly Wages." *Journal of Labor Economics* 8(1): S329-S363.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2003. Fifty Years of Mincer Earnings Regressions. Working Paper no. 9732, National Bureau of Economic Research, Cambridge, MA.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. Wages, Productivity and Worker Characteristics: Evidence from Plant – Level Production Functions and Wage Equations. *Journal of Labor Economics* 17, no. 3: 409 - 446.

Hellerstein, Judith K., David Neumark. 2004. Production Function and Wage Equation Estimation with Heterogeneous Labor: Evidence from a new Matched Employer – Employee Data Set. Working Paper no. 10325, National Bureau of Economic Research, Cambridge, MA.

Ilmakunnas, Pekka, and Mika Maliranta. 2005. Technology, Labour Characteristics and Wage–productivity Gaps. *Oxford Bulletin of Economics and Statistics* 67, no. 5: 623-645.

Israeli, Osnat. 2007. A Shapley-based Decomposition of the R-Square of a Linear Regression. *Journal of Economic Inequality* 5:199–212.

Jones, Patricia. 2001. Are educated Workers Really more Productive? *Journal of Development Economics* 64:57-79.

Keane, P. Michael, Kenneth I. Wolpin. 1997. "Career Decisions of Young Men". *Journal of Political Economy* 105(31): 473-522.

Lallemand, Thierry, Robert Plasman and François Rycx. 2005. Why do Large Firms Pay Higher Wages? Evidence from Matched Worker–firm Data. *International Journal of Manpower* 26, no. 7/8: 705 – 723.

Leontaridi, Rannia Marianthi. 1998. Segmented Labour Markets: Theory and Evidence. *Journal of Economic Surveys* 12(1): 63-101.

Martins, Pedro. 2008. Dispersion in Wage Premiums and Firm Performance. Working Paper no. 8, Centre for Globalization Research, London.

Menezes-Filho, Naercio A., Marc-Andreas Muendler, and Garey Ramey. 2006. The Structure of Worker Compensation in Brazil, with a Comparison to France and the United States. University of São Paulo, University of California, San Diego.

Mercado, Alejandro F., Lykke E. Andersen and Beatriz Muriel H. 2003. Discriminación Étnica en el Sistema Educativo y en el Mercado de Trabajo de Bolivia, *Revista Latinoamericana de Desarrollo Económico* 1 (September): 69-98.

Mercado Alejandro F., and Tirza J. Aguilar. 2006. Market and Inequality Revisited. Working Paper no. 1, Institute of Socio-Economic Research (IISEC), Catholic University of Bolivia, La Paz, Bolivia.

Mincer, Jacob A. 1974. Schooling, Experience and Earnings. New York: NBER Press.

Moensted, Trine. 2000. Wage Differentials between the Formal and the Informal Sector in Bolivia. Working Paper no. 2, Institute of Socio-Economic Research (IISEC), Catholic University of Bolivia, La Paz, Bolivia.

Morduch, Jonathan, and Terry Sicular. 2002. Rethinking Inequality Decomposition, with Evidence from Rural China, *The Economic Journal* 112 (January): 93-106.

Moretti, Enrico. 2004. Workers' Education, Spillovers and Productivity: Evidence from Plant – Level Production Functions. *American Economic Review* 94, no. 3: 656-690.

Muravyev, Alexander. 2009. Employer size, Wages and Unobserved Skills: Evidence from Moonlighters in the UK. *The Manchesteer School* 77, no. 6: 651–674.

Muriel, Beatriz H. 2005. Female Labor Market Conditions in Urban Bolivia. Working Paper no. 3, Masters for Development, Catholic University of Bolivia, La Paz, Bolivia.

Muriel, Beatriz H. 2011. Rethinking Earnings Determinants in the Urban Areas of Bolivia. Working Paper no. 6, Institute for Advanced Development Studies (INESAD), La Paz, Bolivia.

Muriel, Beatriz H., and Luis Carlos Jemio. 2010. Mercado Laboral y Reformas en Bolivia. In *Desafíos laborales en América Latina después de dos décadas de reformas estructurales. Bolivia, Paraguay y Perú 1997-2008*, ed. José Rodríguez and Albert Berry. Lima, Peru: Instituto de Estudios Peruanos.

Gürtzgen, Nicole. 2009. "Firm Heterogeneity and Wages Under Different Bargaining Regimes: Does a Centralised Union Care for Low-productivity Firms?". ZEW Discussion Papers no. 08-130, Leibniz Information Centre for Economics.

Oi, Walter Y., and Todd L. Idson. 1999. Firm-Size and Wages. In *Handbook of Labor Economics* (volume 3B), ed. Orley C. Ashenfelter and David Card. North-Holland: Elsevier.

Pérez de Rada, Ernesto. 1997. Discriminación Salarial por Género y Etnia en Ciudades Principales de Bolivia, Working Paper no. 47, Unidad de Análisis de Políticas Sociales, La Paz-Bolivia.

Ramírez, Rocio. 2003. Pobreza y Discriminación Laboral por Género. Graduate Department of Economics, Catholic University of Bolivia, La Paz, Bolivia.

Reza Arabsheibani G. and Altay Mussurov. 2006. Returns to Schooling in Kazakhstan: OLS and Instrumental Variables Approach. Discussion Paper no. 2462 (November), The Institute for the Study of Labor (IZA), Bonn, Germany.

Rivero, Roberto, and Wilson Jiménez. 1999. Diferencias Salariales en el Mercado de Trabajo Urbano en Bolivia, 1981-1997, *Revista de Análisis Económico* 17: 15-45.

Soderbom, Mass, Francis Teal and Anthony Wambugu. 2005. Unobserved Heterogeneity and the Relation between Earnings and Firm Size: Evidence from two Developing Countries, *Economics Letters* 87:153 – 159.

Shorrocks, Anthony F. 1982. Inequality Decomposition by Factor Components. *Econometrica* 50, no. 1: 193-211.

Spatz, Julius, and Susan Steiner. 2002. Post-Reform Trends in Wage Inequality: The Case of Urban Bolivia. Working Paper no. 1126, Kiel Institute for World Economics, Kiel, Germany.

Sullivan, Paul. 2007. "Estimation of an Occupational Choice Model when Occupations are Misclassified." *The Journal of Human Resources* 44(2): 495-535.

Temesgen, Tilahun. 2005. Determinants of Wage Structure and Returns to Education in a Developing Country: Evidence from Linked Employer – Employee Manufacturing Survey Data of Ethiopia. *Seoul Journal of Economics* 18, no. 4: 277-203.

Trefler, Daniel. 1993. International Factor Price Differences: Leontief was Right. *Journal of Political Economy* 101, no. 6: 961-87.

Troske, Kenneth R. 1995. The Worker – Establishment Characteristics Database. Center for Economic Studies, Bureau of the Census, Washington, DC.

Verner, Dorte. 1999. Are Wages and Productivity in Zimbabwe Affected by Human Capital Investment and International Trade? Policy Research Working Paper no. 2101, World Bank, Washington, DC.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MA MIT Press.

Woodcock, Simon D. 2008. Wage Differentials in the Presence of Unobserved Worker, Firm and Match Heterogeneity. *Labour Economics* 15: 772-794.

Yasar, Mahmut, and Catherine J. Morrison. 2008. Capital–skill Complementarity, Productivity and Wages: Evidence from Plant–level Data for a Developing Country", *Labour Economics* 15:1–17.

Van Biesebroeck Johannes. 2011. Wages Equal Productivity. Fact or Fiction? Evidence from Sub Saharan Africa. *World Development* 39(8): 1333–1346.

van Ours Jan and Lenny Stoeldraijer (2011), "Age, Wage and Productivity in Dutch Manufacturing," De Economist 159(2): 113–137.