An Investigation of the Labour Market Earnings in Deprived Areas: A test of labour market segmentation in the Slums

João Pedro Azevedo
IPEA
University of Newcastle

jazevedo@ipea.gov.br

DRAFT

Comments welcomed

Resumo

Este artigo investiga os determinantes dos retornos salariais de homens e mulheres que residem em comunidades de baixa renda da cidade do Rio de Janeiro. Os dados utilizados foram a Pesquisa Comunidades de Baixa Renda, que visitou 21,704 domicílios em 51 comunidades de baixa renda na cidade do Rio, e dados georeferencializados gerados a partir do Censo de 2000. O trabalho sugere que a opção pelo mercado informal pode ser uma escolha racional, do ponto de vista do trabalhador, em particular enquanto jovem. O trabalho também sugere a existência de efeitos de seleção, os quais evidenciam a não aleatoriedade do processo de escolha ocupacional, os quais são corrigidos através da adoção métodos de correção em dois estágios.

Palavra-chave: pobreza urbana, segmentação, mercado de trabalho, escolha ocupactional, favelas
Tema : Setor informal; Diferencial salarial; Escolha ocupacional
Código JEL: J24, J31, J42, O15, 017
AREA 12: Economia do Trabalho

Abstract

This paper presents empirical evidence on the determinants of labour market earnings for males and females in the slums of Rio de Janeiro. The data used is from a survey of 21,704 households at 51 slums from the city of Rio de Janeiro, augmented with contextual information from the 2000 census as well as GIS information on the location of these communities. Several tests of the labour market segmentation hypothesis were implemented. This work suggests that the informal sector can be a rational choice from the workers point of view, in particular at a younger age. In addition, it also supports the existence of selection effects, which suggests that occupational choices are non-random, and must be taken into consideration through a two-step estimation of the earnings equation.

keywords: urban poverty, segmentation, labour markets, occupational choice, slums
Theme : Informal Sector; Wage Gap; Occupational Choice
JEL-Code: J24, J31, J42, O15, 017
AREA 12: Labour Economics

1 Introduction

In most developing countries the incidence of poverty is higher in the rural areas, however, the population in those countries is urbanizing quite rapidly (Ravallion, 2002). Moreover, in some regions of the planet, such as Latin America, the urbanization of poverty is already a fact. Although poverty rates are higher in rural areas than in urban areas in the region (55% vs. 39%), in absolute terms there are more than twice as many urban poor than rural poor - 68 million rural poor compared with 138 million urban poor (Cira, 2002).

Most of the times, the urban poor is only able to afford housing in areas characterized by precarious public health, a lack of basic infrastructure (including water delivery), inadequate public services, substandard sanitation, and widespread violence and insecurity, such areas are also referred as slums. According to a new report from the United Nations Human Settlements Program, "The Challenge of Slums: Global Report on Human Settlements 2003", at least 1 billion people worldwide currently live in urban slums (UN-HABITAT, 2003). Within 30 years, if current trends continue, one out of every three people on the planet will live in such conditions.

It is clear then, that urban poverty and the dynamics of the slums have a substantial importance on the development debate.

For a long time, studies on economic development have emphasized the agriculture and industry sector, following the prevailing paradigms of Lewis (1954) and Fei and Ranis (1964), thus underestimating the importance of the entrepreneurial sector. However, the work by Harris and Todaro (1970) illustrated that migrants from agriculture to industry may face a period of unemployment or, if unemployment is unaffordable, may be forced to provide themselves through a low-productivity household enterprise. The latter idea took more shape when an ILO (1972) report defined the so-called informal sector, which has proved to be one of the most influential concepts in development economics.

More often than not, in the most deprived areas of developing countries, entrepreneurship and informal sector activities are, in spite of the numerous hurdles that they have to overcome, an important port of entry in the labour market (De Soto, 1989, De Soto; Gill et al., 1999). In this context, it is extremely relevant to understand the labour market of these areas, and identify ways to support the entrepreneurial initiatives of this segment, both as an employment and income generation strategy.

This paper intends to contributes to a virtually non existent literature on the labour market from extremely deprived urban areas in developing countries, with a particular interest on understanding the wage differential of the residents in these communities and how it is related to the entrepreneurial activity within these communities.

Most of the previous work comparing wages in developing countries analyze occupation as a binary choice, either formal or informal, or wage earners and self-employed (see Van Der Sluis et al. (2003) for a recent survey of this literature). One of the problem in these studies is the assumption that some categories can be pooled together, or the exclusion of some important categories such as the self-employed or the wage earner. In this paper, we follow a smaller literature that models the multinomial nature of these choices (Saavedra and Chong, 1999; Cohen and House, 1996). Moreover, this paper shows through an often neglected specification test of the multinomial logit model that these categories can not be pooled together, and applies a relatively recent polynomial selection bias correction model (Bourguignon et al., 2001) which, to the best of the author's knowledge, has never been applied in this literature.

Another problem in the existing literature is its often narrow focus on the gender dimension, either by pooling males and females together, or by analyzing only one gender. This work tries to provide a more complete picture of the labour market under scrutiny, analyzing the labour market conditions of both males and females separately.

Many of the contributions of this paper have only been made possible given the unique character of the data set used, which consist of the interviews 21,704 household from 51 different slums from the city of Rio de Janeiro. This data set has been augmented with contextual information from the 2000

census data at the neighborhood level from which each community is located, as well as geographical information on the position from each of these communities in the city of Rio de Janeiro.

The present work is structured in five sections, including this introduction. The next section presents a review on the literature on earnings and occupational choice, and how this models are often estimated. The third section describes the data set used in this study. The following section presents the empirical results of the study with several tests for the geographical and labour market segmentation for the residents of the slums of Rio de Janeiro. The fifth and last section highlights the major findings of this work.

2 Models of Labour Earnings and Occupational Choice and Their Estimation

Various models of labour market earnings and occupational choice have been advanced in the economic literature. Mincer (1974, 1958) and Becker (1964) human capital models represent one of the most prominent representatives of this literature.

The simple schooling model of Mincer (1958) is given by:

$$ln(y_i) = \beta_0 + \beta_1 s_i + \mu_i, \tag{1}$$

where y_i is earnings, s_i is schooling, and μ_i is an error term for the model representing unobserved determinants of earnings. β_1 can be interpreted as the average return to schooling with β_0 as the logarithm of the base earnings on an individual with no schooling.

In latter works, like Mincer (1974) and Becker and Chiswick (1966), Equation (1) is augmented to include the effect of investment after formal schooling. In this specification the log of earnings is written as

$$ln(y_i) = \beta_0 + \beta_1 s_i + \beta_2 t_i + \beta_3 t_i^2 + \mu_i^*, \tag{2}$$

where t_i and t_i^2 are, respectively, post-schooling labour market experience and experience squared, and μ_i^* is an unobserved disturbance term. Often this basic equation is enlarged, including hours of work, regional dummy variables, and other variables (see Mincer (1974)). Model (2) is the basic specification of the earnings equations adopted in this study.

Another important extension of Model (2) is when the sample is divided in J sub samples and separate earning equations are estimated for each segment. Models such as this are often used when separate earnings equations are estimated for different gender, occupational and geographical groups. This model can be formally presented as follows:

$$ln(y_{ij}) = \beta_{0j} + \beta_{1j} s_{ij} + \beta_{2j} t_{ij} + \beta_{3j} t_{ij}^2 + \mu_{ij}^*, \tag{3}$$

where the subscript j indicates which sub sample this earnings equation refers to.

Model (3) will provide unbiased estimates of (β_k)'s as long as the the population is randomly distributed across the j segments. In other words, this model will no longer holds if individuals are heterogeneous with respect to: their preferences; their observed and unobserved productivity-enhancing characteristics; and, the strength of their contacts in the labour market. Nevertheless, when the sample separation is known, combining a set of observable characteristics for each individual and the knowledge of the j segment in which individuals are located provides both an insight into the nature of the covariance and a means to construct unbiased estimates of the parameter vectors . Hence, to achieve this goal, it is necessary to model also the occupational choice process.

In order to combine the equation of earnings determination with occupational choice (omitting the individual-specific subscript i and j is a categorical variable that describes the choice of an economic agent among J alternatives), we can follow Bourguignon et al. (2001) notation:

$$y_j = x_j \beta_j + \mu_j^* \tag{4}$$

$$y_j^* = z_j\gamma_j + \eta_j, \qquad j = 1 \ldots J \tag{5}$$

where all the variables x_j and z_j, are exogenous, and the disturbance μ_j^* verifies $E(\mu_j^*|x,z) = 0$ and $V(\mu_j^*|x,z) = \sigma_j^2$. The outcome variable y_j is observed if, and only if, the category j is chosen, which happens when

$$y_j^* > \underset{s \neq j}{max}(y_s^*) \tag{6}$$

Condition (6) can also be written as:

$$z_j\gamma_j > \varepsilon_j, \tag{7}$$

where

$$\varepsilon_j = \underset{s \neq j}{max}(y_s^* - \eta_j) \tag{8}$$

Now it is possible to assume that the (η_j)'s are independent and follows, identically, a Gumble distribution, with cumulative and density functions as, respectively: $G(\eta) = exp(e^{-\eta})$ and $g(\eta) = exp(-\eta - e^{\eta 0})$. This specification leads to the familiar multinomial logit model, presented as:

$$P(z_j\gamma_j > \varepsilon_j) = \frac{exp(z_j\gamma_j)}{\sum_s exp(z_s\gamma_s)} \tag{9}$$

where maximum likelihood estimates of the (γ_s)'s can be easily obtained. Note that Equation (9) represents the probability that an individual with characteristics z_j chooses the segment j. Therefore, this can be taken as the empirical estimation of the occupational choice process. [1]. Often, vector z_j contains variables such as education [2], age [3], and marital status [4].

In order to estimate unbiased β_j's, Heckman (1979), in a seminal article, proposes a two stage technique to correct for selection bias in the case of a dichotomous choice model ($J = 2$). A few authors have extended this work and proposed similar two-stage correction procedures for polychotomous choice models, for instance Dubin and McFadden (1984), Lee (1983), and Bourguignon et al. (2001).

A paper by Schmertmann (1994) demonstrates that when the number of potential outcomes

[1] Models of occupational choice and entrepreneurship have also received great attention by economists, often taking self-employment as a working definition of entrepreneurship (see Parker (2004) for a comprehensive survey on this subject). The development literature has also produced a family of models to shed light on the process of occupational choice, income inequality and economic development (for more details see Banerjee and Newman (1993) and Ghatak and Jiang (2002)).

[2] The role of education qualifications has been incorporated into many empirical studies, being a key determinant of success in the labour market. Educational attainment may act as a proxy for ability - individuals of higher ability many make better managers, which in turn may enhance their probability of becoming self-employed. Van Der Sluis et al. (2003) have recently produced a comprehensive survey of the effects of education on entrepreneurship in developing countries. Studies reporting a positive relationship between education attainment and the probability of self-employment, include (Rees and Shah, 1986; Borjas, 1986; Borjas and Bronars, 1989; Evans and Leighton, 1989). Alternatively, higher levels of educational attainment may play a signalling role in the labour market with high educational qualifications serving to secure employment in the non self-employment sector. Evidence supporting this inverse relationship between higher educational qualifications and the propensity to become self-employed include Evans (1989) and De Wit and van Winden (1989). In summary, the evidence regarding the relationship between education and the propensity to become self-employed remains inconclusive.

[3] An individual's age may also affect his/her propensity to become self-employed via a number of different channels (Lucas, 1978; Calvo and Wellisz, 1980). For instance, age may act as a proxy to capture the effect of an individual's awareness, knowledge and experience in the labour market, thereby reflecting general human capital. Alternatively, as an individual becomes older, he/she may have accumulated the financial resources required for self-employment; hence age may capture effects related to financial, as well as human, capital.

[4] Often, marital status has been incorporated into many empirical studies, since it is assumed to represent stability and, as such, may provide a suitable background for 'risk' self-employment. Moreover, Bernhardt (1994) and Blanchflower and Oswald (1990) find that having a working spouse enhances the probability of self-employment. Similarly, Schiller and Crewson (1997) find evidence of intra-coupling risk pooling with a husband's primary employment increasing the probability that a wife will be observed in self-employment.

exceeds two ($J > 2$), the Lee technique contains implicit and unnecessary restrictions on the structure of the error terms that generally render it inappropriate for a study of this sort. Thus, a few authors, such as Cohen and House (1996), have argued that the Dubin and McFadden (1984) procedure was the most adequate. In this alternative procedure, each segment earnings equation also includes a separate selection term for each pairwise utility comparison, which sets for a J sector model $J-1$ selection terms.

However, Bourguignon et al. (2001) propose an alternative selective correction method that involves all correlation coefficients between the disturbance term of the outcome equation of interest (Equation (5)) and the disturbance term of all categorical latent expression (Equation (4)). Thus, in a J sector model there are J selection terms.

This paper applies this recent contribution, and use the following equation to estimate unbiased β_j's:

$$y_1 = x_1\beta_1 - \sigma_1\left[\tilde{\rho}_1 m(P_1) + \sum_{j>1}\tilde{\rho}_j \frac{P_j}{(P_j-1)} m(P_j)\right] + \nu_j \tag{10}$$

where ν_j is orthogonal to all other terms in the RHS and has zero expectation. Because of this property, least squares may now be used to estimate the β_1's. It is important to notice that $(\sigma_1\tilde{\rho}_1), \ldots, (\sigma_1\tilde{\rho}_j)$ are the coefficients for the correction terms which adjust for selection bias, and $(m(\hat{P}_j))$'s are the fitted probabilities of Model (9). Henceforth, we refer to this model as the BFG.

Overall, the estimation procedures in this paper will consist of the following procedures. First, according to the tradition of a well-established empirical literature, we will estimate embellished variations of Equations (2) and (3) by ordinary least squares (OLS). Second, Model (10) is estimated through a two-step procedure: (i) estimate the multinomial logit (Model (9)), and derive from it the fitted probabilities $(\hat{P}_j)$'s using the $(\hat{\gamma}_j)$'s [5]; and, (ii) estimate Equation (10) by least squares.

3 Data and Summary Statistics

The data for this study was obtained from a survey carried out by the municipality of Rio de Janeiro between January/1998 and March/2000. The survey covered households and enterprises in 51 out of the 462 slums in the city of Rio de Janeiro at the time of the survey[6]. The survey had two modules which were carried out independently: household (PCBR-Dom) and enterprises (PCBR-Est). The data from both modules was collected through face-to-face interviews conducted by enumerators and field supervisors trained by the Brazilian National School of Statistics (ENCE) which is part of the Brazilian National Office of Statistics (IBGE). The households module followed a questionnaire similar to the World Bank LSMS[7] and covered a sample of 21,704 households in the 51 slums selected. The enterprise module is a census of all the 4,553 business establishments located in these slums. This was one the most extensive efforts to map these communities. A full description of the survey is contained in Silva et al. (1998).

The current study only uses the data of the household module (PCBR-Dom).

[5] Of course, the $(\hat{P}_j)$'s and $(m(\hat{P}_j))$'s depends on all (z_j)'s and incorporated all the information behind the multinomial logit model. Their combination on the RHS of (10) yields a consistent estimates of the conditional expected value of the residual term in the original outcome equation (4).

[6] According to the 1991 Census, there are 462 slums or *favelas* in city of Rio de Janeiro, with a total population of 882,667 in 224,350 households (Amsberg, 1999), or 16% of the total population of the city of Rio de Janeiro.The data from the 2000 census only seems to reinforce this trend. Moreover, the population living on this condition has grown 23% from 1980 to 1991, which represents a threefold increase in comparison to the citywide population growth average.

[7] The Living Standards Measurement Study (LSMS) was established by the World Bank in 1980 to explore ways of improving the type and quality of household data collected by government statistical offices in developing countries. The objectives of the LSMS were to develop new methods for monitoring progress in raising levels of living, to identify the consequences for households of current and proposed government policies, and to improve communications between survey statisticians, analysts, and policymakers.

>From the 21,704 households covered by the survey, we accessed information on 82,948 individuals (who live in the same house as the respondent). From these 82,948 individuals, we restricted our analyzes according to five criteria:

a) all men and women between the ages 15 to 65, inclusive. This choice reduced our sample to 52,514 individuals.

b) individuals who were working on the week of the survey, which restricted our analysis to 29,320 individuals. We removed from our sample 4,169 students, 2,965 early retirement, 3,986 unemployed, 9,551 unpaid/domestic workers and approximately 2,500 classified as others (see Table (1)).

Tabela 1: **Activity During the Reference Week (Age 15 to 65))**

Item	Number	Per cent
Worked	29,320	56
Had work but did not work	327	1
Looked for Job	3,986	8
Early Retired	2,965	6
Student	4,169	8
Domestic	9,551	18
Other	2,196	4
Total	52,514	100

Source: PCBR-Dom

c) individuals with only one occupation on the week of reference; this restricted the sample to 28,930 out of the 29,320, workers.

d) in the household survey, there were six (6) occupational segments :wage earners; domestic workers, cooperative members, employers, self-employed, and unpaid (see Table 2). We also excluded the segments of domestic workers[8], unpaid workers, unemployed, and cooperatives members.

Tabela 2: **Occupation Distribution (Age 15 to 65 and only one occupation)**

Item	Number	Per cent
Wage Earner	19,579	68
Domestic Worker	3,799	13
Cooperative members	161	1
Employer	353	1
Self-employed	4,804	17
Unpaid	234	1
Total	28,930	100

Source: PCBR-Dom

e) the sample is further reduced by the absence of data on other covariates, including earnings, used in the empirical analysis.

[8]Following common practice, we excluded domestic servants segment, since they receive an unknown portion of their pay in kind (see (Telles, 1993, page 234) and (Marcouiller et al., 1997, page 3))

The final sample then comprises 19,332 individuals, of whom 13,608 (70.4%) are male.

Due to the time differences of the data collection, the consumer price index of the metropolitan region of Rio de Janeiro's city was used to deflate all the monetary values in the survey to prices of March of 2000.

Most of the variables were computed directly from the replies to the survey questionnaire. One notable exception is OCCUPFOR, which was generated by crossing the information from occupational choice with the data on social security enrolment and enterprize location. Formal and informal wage earners were classified according to their social security status: workers with no social security were classified as informal, while workers with social security were classified as formal. In this paper we follow the standard literature, and take self-employment and employer categories as our working definition of entrepreneurs (see Parker (2004) and Le (1999)).The ideal variable to define entrepreneurs as formal or informal is the business registration; however the survey used did not carry this question. In order to overcome this limitation, we follow Magnac (1991) and adopted a definition of formality and informality based on the business infrastructure. Entrepreneurs which operated their enterprises from home, were classified as informal, while entrepreneurs that operate their business from a separate address were classified as formal.

Table 3 shows the sample distribution of OCCUPFOR by gender. A couple of interesting features of this table are worth highlighting. First, most of our final sample comprises of formal wage earners (11,419). Second, although the gender distribution if very similar on the formal sector, with around 70% males on both wage earners and entrepreneurs, females tend to dominate, on the informal sector, the entrepreneur category, with 57% of this occupation.

Tabela 3: **Occupation and Formality by Gender**

	Gender		
OCCUPFOR	male	female	Total
Wage (Formal)	7,993	3,426	11,419
	70	30	100
Wage (Informal)	3,276	977	4,253
	77	23	100
Entrepreneur (Formal)	2,019	900	2,919
	69	31	100
Entrepreneur (Informal)	320	421	741
	43	57	100
Total	13,608	5,724	19,332
	70	30	100

Source: PCBR-Dom

A few aspects of the data need to be borne in mind when the present results are analyzed and compared with results from other studies. It is important to recognize that the data is not statistically representative of the entire population that reside in slum in the city of Rio. The selection criteria used to include each community in the survey sample was its involvement in a housing and infrastructure program from the municipality of Rio (for more details see Riley et al. (2001) and AciolyJr. (2001)). The 1998/2000 phase of this program prioritized small and medium size slums with already some infrastructure, leaving the most challenging and destitute communities out. If anything, the existing anecdotal evidence seems to suggests that these communities are among the least problematic, and should be interpreted as an upper boundary of the reality of slums in Rio de Janeiro.

This study also uses two complementary data sets, namely the 2000 Census (IBGE, 2000) and the 1996 geographical boundaries of the slums and neighborhoods of the city of Rio de Janeiro (LabGeo, 1996b,a).

The 2000 Census data was used to generate average incomes for each neighborhood where the analyzed slums are located. The goal of this variable is to capture geographical effects that might

arise from slums nearby more affluent neighborhoods.

The data from the 1996 geographical boundaries of the slums and the city of Rio de Janeiro was used to generate distance measures from the centroids of each slum to downtown Rio, in order to control for transportation effects that might affect the labour market of each of these communities.

4 Empirical Analysis

This section analyzes the wage differential among males and females workers working on both the formal and informal sector, as either wage earners or entrepreneurs.

4.1 Labour Market Segmentation

While the previous exercise showed a strong geographical segmentation, this section of the study aims to further investigate the possibility of labour market segmentation in the slums of Rio de Janeiro. In order to do that, we first briefly present a simple stylized model of a dualistic labour market. We then move the empirical estimation of a few models that gradually shed light on this issue.

Dualistic or segmented models of labour market have pervaded the economic development literature since the seminal work of Lewis (1954) (see Cain (1976) for a survey of dual market theory). According to this hypothesis, the existence of dual labour markets will be reflected in dualism in earnings as long as workers with similar characteristics are paid different wages given the sector in which they work (Saavedra and Chong, 1999).

The wage segmentation hypothesis can be formally simplified. To do this, consider an economy populated by agents who differ in terms of a finite list X of personal characteristics. They are employed either in the formal (F) sector or the informal (I) sector. Both sectors offer a menu of jobs described by a vector Y of characteristics that include industry and establishment size.

Let $w^F(X, Y, \varepsilon)$ and $w^I(X, Y, \varepsilon)$ denote integrable random variables that give the agent's log earnings in, respectively, the formal and the informal sector, as a function of their personal and job characteristics, and an exogenous source of uncertainties denoted by ε. The wage segmentation hypothesis can be stated as:

$$S : E(w^F(X, Y, \varepsilon) - w^I(X, Y, \varepsilon)|X, Y \in A) > A$$

for a non-negligible subset A or characteristics.

In this section, we ask weather such a subset of personal and job characteristics can be found in the set of workers sampled in the slums of Rio de Janeiro.

One way of assessing whether or not there are differences in income determination is by comparing income differentials of individuals that are observationally equivalent. If they persist, it may be because such differentials are not able to be explained by the differences of the observable characteristics of the individuals, but by the characteristics of the job.

We can now estimate a variant of Model (2) in which monthly earnings, y, is assumed to be log-linear, and the controls used are: - dummy variables created from OCCUPFOR - D_{wi}, D_{of} and D_{ow} - that indicate whether or not the individual is a wage earner from the informal sector, an entrepreneur that belongs to the formal sector, or an entrepreneur from the informal sector [9]; years of schooling; experience; gender; marital status; migrant status; occupational sector; number of hours per week; job location; neighborhood average income; and, average distance from the community to downtown Rio.

Earning differentials in the distinct occupational categories are obtained from the parameters γ_1, γ_2 and γ_3, which respectively measure the effects of variables D_{wi}, D_{of} and D_{ow}. The results of the regressions are show in Table 4.

[9] The fourth possible category, that is, wage earner from the formal sector, is used as the control category

The coefficients of Table 4 suggest that entrepreneurs have higher returns than wage earners. Even after controlling for the above mentioned covariates, the parameters for formal entrepreneurs, informal entrepreneurs and informal wage earners remained positive and statistically significant. This suggests that these occupations have a higher return than the reference category, namely, formal wage earners. Moreover, the gender coefficients suggests that the controls do explain part of the earning differential between men and women, given that this differential has now dropped from 30% to 16%, nevertheless, it still present and significative.

A more general comment on findings of Table 4 is the already noted much lower return to education of the residents of this communities, when compared with the existing empirical evidence for developing countries. Another point is the average 19% wage increase, for occupations outside the slums.

Tabela 4: **Earnings Differential Between Formal and Informal Workers**

	lnwage
	(1)
Education (Yrs of Schooling)	.003
	(.0007)***
Education Sqr	.005
	(.00005)***
Experience	.030
	(.0002)***
Experience Sqr	-.0005
	(3.38e-06)***
Female	-.214
	(.001)***
Marital statua (married)	.076
	(.001)***
Migrant	.066
	(.001)***
Head of Household	.087
	(.002)***
Neighborhood Average Income	.0001
	(1.71e-06)***
Slum average distance to Downtown Rio	3.85e-07
	(6.46e-08)***
Works inside the community	-.193
	(.002)***
γ_2	-.036
	(.001)***
γ_3	.168
	(.002)***
γ_4	-.089
	(.004)***
Obs	728194.000
R^2	.287
F statistic	12390.020

Standard errors in parentheses; Significance levels::10% **:5% ***:1%.*
Note:Sector and Time Dummies have been omitted
Source:PCBR-Dom

Using an index of raw average wages, in which again the formal wage earners are the base category, we further investigate the wage differentials. The first panel of Table 5 shows that a formal entrepreneur earn 22% more than a formal wage earner. On the other hand, it is found that the informal entrepreneurs earn, on average, only 2% more than formal wage earner.

The second panel of Table 5 shows an index of monthly earnings, obtained from the parameters of the regression presented in Table 4, that captures the predicted earnings of individuals, whose only difference are there non-observable characteristics. Earning differentials among categories are reduced dramatically when controls are introduced. For both males and females wage earners, earnings differentials between formal and informal workers become unimportant. However, earning differential among the entrepreneurs remain significant, with those entrepreneurs working on the formal sector earning 15% more and those working on the informal sector earning 8% less. In summary, this exercise shows that, when controlling for observable, including occupation, entrepreneurship is the best occupation for males, and the worse one for females.

There are a few possible explanations why controls for education and experience do not completely reduce earnings differentials for the entrepreneurs. As Saavedra and Chong (1999) discussed, this empirical specification assumes only parallel shifts in the earning structure, but does not consider that the full earnings generating process may be different for the different categories. As Lazear (2003) argues, entrepreneurs are "jacks-of-all-trades", in other words, individuals with a broad set of abilities, which can not all reduced to education and experience. Also, an implicit assumption that individuals are randomly distributed across sectors may be unrealistic. On the other hand, neither all the factors that determine the sectoral allocation of workers, nor the earnings structure may be observable. For instance, there may be unobservable differences in terms of quality of education or innate ability of the individuals that may determine self-selection of individuals to specific sectors.

Tabela 5: **Indices of Monthly Earnings**

	All	Male	Female
I. Index of Raw Averages			
Wage Earner			
Formal Sector	100	100	100
Informal Sector	95	93	93
Entrepreneur			
Formal Sector	123	127	110
Informal Sector	81	100	78
*II. Index of Predicted Monthly Earnings**			
Wage Earner			
Formal Sector	100	100	100
Informal Sector	90	88	90
Entrepreneur			
Formal Sector	111	116	99
Informal Sector	67	76	72

* Obtained from coefficientes of Table 4 evaluated at the sample mean.

Source: Author's analysis of PCBR-Dom.

An alternative approach to test the labour market segmentation hypothesis is to analyze differences in the earnings generating process, by estimating earnings equations for each sector separately (Marcouiller et al., 1997; Telles, 1993). In brief, the labour market is divided into 4 sub samples (one for each potential segment: formal wage earners; informal wager earners; formal entrepreneurs; and, informal entrepreneurs) and separate earnings equations are estimated for each segment using OLS. The hypothesis of labour market segmentation is tested by conducting a series of pairwise F-tests to determine whether the estimated coefficients vary across equations.

The form of our earning equation follows the standard human capital model. Controlling for

sector, marital status, job location, exogenous neighbourhood factors and migration status, we analyze if the the returns to labour differ systematically between the distinct occupational categories for observationally identical workers.

Following Equation (3), earnings functions are estimated separately for men and for women in each of the occupational categories (OCCUPFOR). The coefficients for both men and women have expected signs, and are presented on Table 6 and Table 7.

The overall picture that emerge is that the difference between wage earners and entrepreneurs is much greater than the difference between formal and informal sector workers. We also find statistically significant formal wage premia for both education and experience, as well as higher returns for those who manage to work outside the slums. However, we find that females informal sector workers enjoy a higher return than their counterparts in the formal segment (both wager earners and entrepreneurs - see table 7).

Tabela 6: Segmented Wage Equation Male Workers: OLS

	Wage Earner				**Entrepreneur**			
	Formal		**Informal**		**Formal**		**Informal**	
		Robust		**Robust**		**Robust**		**Robust**
Variable	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
Education	0.011**	0.00	0.002	0.00	0.018**	0.00	0.063**	0.01
Education Sqr	0.004**	0.00	0.005**	0.00	0.003**	0.00	-0.003**	0.00
Experience	0.035**	0.00	0.038**	0.00	0.018**	0.00	0.025**	0.00
Experience Sqr	-0.001**	0.00	-0.001**	0.00	-0.000**	0.00	-0.001**	0.00
Marital statua (married)	0.043**	0.00	0.087**	0.01	0.062**	0.01	-0.030*	0.02
Migrant	0.060**	0.00	0.100**	0.00	0.120**	0.00	0.145**	0.01
Head of Household	0.094**	0.00	0.069**	0.01	0.176**	0.01	0.215**	0.02
Commerce	-0.116**	0.00	-0.376**	0.00	-0.377**	0.01	-0.389**	0.02
Service	-0.122**	0.00	-0.288**	0.00	-0.239**	0.01	-0.207**	0.02
Construction	-0.007*	0.00	-0.196**	0.00	-0.036**	0.01	-0.002	0.02
Transport	0.073**	0.00	-0.111**	0.01	0.240**	0.01	0.234**	0.03
Hours per week (+20hrs to 40hrs)	0.282**	0.02	0.400**	0.01	0.483**	0.02	0.447**	0.04
Hours per week (+40hrs to 44hrs)	0.281**	0.02	0.461**	0.01	0.551**	0.02	0.490**	0.05
Hours per week (more than 44hrs)	0.335**	0.02	0.503**	0.01	0.652**	0.02	0.685**	0.04
Neighborhood Average Income	0.000**	0.00	0.000**	0.00	0.000**	0.00	0.000**	0.00
Slum average distance to Downtown Rio	0.000*	0.00	0.000**	0.00	-0.000**	0.00	-0.000**	0.00
Works inside the community	-0.128**	0.00	-0.233**	0.00	-0.115**	0.01	-0.069	0.10
Constant	4.764**	0.02	4.613**	0.01	4.946**	0.02	4.812**	0.11
Number of cases	285094		126154		68288		14866	
Adj. R-Square	0.228		0.321		0.187		0.203	
F-statistic	4592.119		3565.734		921.962		.	
Wald Test								
$F-$value (All)	264.980**							
$F-$value (Wage Earners)	248.789**							
$F-$value (Onw Account)	100.138**							
Significance levels : † : 10% * : 5% ** : 1%								

Tabela 7: Segmented Wage Equation Female Workers: OLS

	Wage Earner				**Entrepreneur**			
	Formal		**Informal**		**Formal**		**Informal**	
		Robust		**Robust**		**Robust**		**Robust**
Variable	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
Education	-0.024**	0.00	-0.026**	0.00	0.015**	0.00	0.040**	0.00
Education Sqr	0.006**	0.00	0.007**	0.00	0.002**	0.00	0.001**	0.00
Experience	0.020**	0.00	0.035**	0.00	0.024**	0.00	0.021**	0.00
Experience Sqr	-0.000**	0.00	-0.001**	0.00	-0.000**	0.00	-0.000**	0.00

Continued on next page...

... table 7 continued

Variable	Coef.	Robust S.E.	Coef.	Robust S.E.	Coef.	Robust S.E.	Coef.	Robust S.E.
Marital statua (married)	0.060**	0.00	0.067**	0.01	0.072**	0.01	0.087**	0.01
Migrant	0.017**	0.00	0.032**	0.01	0.038**	0.01	0.161**	0.01
Head of Household	0.036**	0.00	0.015†	0.01	0.028**	0.01	0.160**	0.01
Commerce	0.005	0.00	-0.119**	0.01	-0.174**	0.03	-0.430**	0.02
Service	-0.079**	0.00	-0.160**	0.01	-0.116**	0.03	-0.185**	0.01
Construction	0.244**	0.02	0.106**	0.03	0.518**	0.03	(dropped)	
Transport	0.181**	0.01	0.188**	0.02	(dropped)		-0.233**	0.02
Hours per week (+20hrs to 40hrs)	0.255**	0.02	0.372**	0.01	0.380**	0.01	0.727**	0.02
Hours per week (+40hrs to 44hrs)	0.266**	0.02	0.381**	0.01	0.380**	0.02	0.655**	0.02
Hours per week (more than 44hrs)	0.279**	0.02	0.395**	0.01	0.545**	0.01	0.736**	0.02
Neighborhood Average Income	0.000**	0.00	0.000**	0.00	0.000**	0.00	0.000**	0.00
Slum average distance to Downtown Rio	-0.000	0.00	-0.000**	0.00	0.000*	0.00	0.000	0.00
Works inside the community	-0.194**	0.01	-0.361**	0.01	-0.062**	0.01	-0.405**	0.08
Constant	4.845**	0.02	4.570**	0.02	4.826**	0.04	4.378**	0.09
Number of cases	132274		42654		31834		27030	
Adj. R-Square	0.199		0.325		0.167		0.200	
$F-$statistic	1559.570		1045.546		433.588		.	
Wald Test								
$F-$value (All)	149.390**							
$F-$value (Wage Earners)	92.463**							
$F-$value (Onw Account)	82.298**							
Significance levels : † : 10% * : 5% ** : 1%								

The F-test to compare the earnings differences between the distinct occupations rejects the possibility that returns are similar across the different sectors (see the last three rows of Tables 6 and 7).

It is important to note that the rejection of the single market hypothesis, like the rejection of the hypothesis of no geographical segmentation, may simply result from sample selection bias (Heckman and Hotz, 1986). As noted in Cain (1976), creating a sample of the basis of the dependent variable of a regression analysis and running separate regressions within each sub sample can produce sharply biased estimated coefficients if in fact a common regression function characterizes all populations. Figure 1 illustrates the point. The solid line in that figure show a hypothetical population earnings equations based on the simple schooling model of Equation (1). The line assumes a positive relationship between income and schooling. The dots around that line represent the residual variability in the earnings equation due to unmeasured variables such as ability and motivation, i.e., the μ_i[10] of Equation (1).

Suppose that there is no segmentation of the market in the sense that the solid line represents the mean earnings function for everyone at all levels of schooling. Creating a sample of low earners (people whose earnings are below C) and running regression on the sample produced biased estimates of the true regression function. As the level of education increases, the proportion of people in the low income category declines. Those who are in the category tend to have low values of μ_i (the error terms representing unobserved determinants of earnings) in Equation (1). As Heckman and Hotz (1986) explain, the measured relationship between income and schooling from samples of low-income workers is biased towards zero because the level of μ_i in the sample of low-income workers is negatively correlated with s_i (years of schooling) even though μ_i is uncorrelated with s_i in the population. The dashed line in Figure 1 indicates the impact of this bias on fitted regression equations. Low-income workers and high-income workers whose behavior is generated by a common regression function (the solid line) appear to be different (the dashed line) only because of the statistical methods used.

Nevertheless, it is possible to correct for this bias using sample selection techniques, which correct for sample selection by accounting for the effect of income truncation (or other selection rules) on the mean of μ in the low income sample. This is the aim of the next section.

[10] error term representing unobservables determinants of earnings

Figura 1: Hypothetical Scatter Diagram and Regression of Earnings on Educational Attainment, With and Without Truncated Earnings.

4.2 Earnings Equation Adjusted for Occupational Selection

So far, we calculated wage premia on the basis of the estimated coefficient of the sectoral earning functions. However, if workers freely choose between the occupation segment, the wage and the sector are simultaneously determined, and the same individual peculiarities which drive one result may also drive the other. In other words, the same unobservable personal characteristics which make an individual more likely to choose the informal sector may also be raising his/her wage in the informal sector above that of a person with the same observed characteristics drawn randomly from the population.

Following Equation (10), presented in Section 2, this paper now estimate the two-step procedures suggested by Bourguignon et al. (2001).

4.2.1 The segment allocation equation and identification

The first step in testing our model of wage determination is to consider a reduced-form selection equation that summarizes what we know about recruitment procedures and participation decision (as described in Model 9). In reduced-form, the choice of a particular segment is a function of segment-specific prices, exogenous income, hours worked, job characteristics, and other individual-specific exogenous variables. The main proxies variables available to capture these effect are: education; age; marital status; partners occupation; a dummy variable indicating if the person is the head of the household; and, a dummy variables indicating if the respondent owns a house.

Table 8 provides the marginal effects of the MNL estimates for the occupation equation. The model proved reasonably successful, converging rapidly with the majority of coefficients significant at the 1% level.

Figures 2, 3 and 4 plots the predicted values of Equation (9). In this particular exercise, all continuous variables were set at its mean values while the categorical variables were set at its most representative figure, married males who own their own house, and whose wife's work as an entrepreneur. It is important to note that in this graphs we are using the variable age as a proxy for experience.

Several interesting findings emerge from these figures. In terms of age (Figure 2), it is clear that workers have an optimal age to be admitted to the formal wage earner category (around 40 years old). Another interesting fact is the crescent importance of experience on the probability of becoming a formal entrepreneur; this corroborates the existing theoretical and empirical literatures on entrepreneurship presented in Section 2. In terms of the informal wage earners, it is clear that the probability of choosing this category decreases as the individual gets older. This fact might suggests that this category plays an important role on the early career of the residents of the poor areas, often being their first job; however, as the individual accumulates experience, they tend to moves away from it, in particular towards entrepreneurial activities. Last but not least, it is clear that experience is not a relevant variable for informal entrepreneurs, having a very small but positive impact on the probability of an individual becoming an entrepreneur.

On the other hand, it is clear that education is an important factor for those individuals working on the formal wage earner category, however for all the other categories its effect is either constant, of negative, as in the case of formal entrepreneurs.

The theoretical literature has also explored the possibility of the effect of wealth on entrepreneurial choice Banerjee and Newman (1993); Evans and Leighton (1989). Using a wealth index (for details on the construction of the index see the appendix), we see that although formal wage earners have substantially more wealth, the only category that present an increasing relationship were the informal entrepreneurs, suggesting that as wealth increases the possibility of becoming and entrepreneur also increases.

Tabela 8: **Marginal Effect: Multinomial Logit**

	Wage Form	Wage Infor	Entr For	Entr Infor
	(1)	(2)	(3)	(4)
Education (Yrs of Schooling)	.009	-.00002	-.008	-.0003
	(.001)***	(.001)	(.0009)***	(.0004)
Age	.011	-.025	.012	.002
	(.002)***	(.002)***	(.002)***	(.0006)**
Age squred	-.00009	.0003	0	0
	(.00003)***	(.00002)***	(.00002)***	(7.71e-06)
Female	.009	-.088	.034	.047
	(.009)	(.007)***	(.007)***	(.005)***
Marital status (married)	-.0009	-.025	.017	.01
	(.009)	(.008)**	(.006)**	(.002)***
Head of Household	.025	-.042	.018	.0002
	(.01)*	(.009)***	(.007)**	(.003)
Partner is an entrepreneur	-.103	.03	.053	.021
	(.016)***	(.014)*	(.011)***	(.005)***
Own a house	-.022	-.0004	.017	.007
	(.012)*	(.01)	(.008)*	(.003)*
Migrant	.031	-.041	.0009	.01
	(.008)***	(.007)***	(.006)	(.003)***
Number of kids at school age	-.004	.0005	.005	-.00009
	(.002)*	(.002)	(.002)**	(.0007)
Number of years residing in the same slum	.0003	-.00009	-.0002	.0002
	(.0003)	(.0003)	(.0002)	(.00008)*
Asset Index	.017	-.012	-.006	.003
	(.003)***	(.002)***	(.002)***	(.0008)***
Asset Index Squared	-.004	.004	.002	-.0004
	(.0007)***	(.0005)***	(.0003)***	(.0002)*

Full Model in the Appendix Significance levels::10% **:5% ***:1%.*

Figura 2: Occupational Choice and Experience.

The picture that emerges from these figures is the fact the entrepreneurs (both formal and informal) are the only categories in with the probability of entering increases over time, suggesting the degree of importance of these occupation have in these communities, and that formal education only has a positive effect for the formal wage earners.

Given the key role of the allocation process in our model, we need to ensure that the model conforms to the assumption of 'independence of irrelevant alternatives' (IIA). This condition is violated when the ratio probability of one alternative to another is affected by the introduction of an additional choice. This amounts to workers expressing differential substitutability and complementarity between alternatives. In such case, the process of sectoral assignment can be better modeled by using a multinomial probit or a nested multinomial logit model.

Testing whether a particular data set adheres to this assumption can be established by using the test suggested by Hausman and McFadden (1984). We systematically compared the four segment model against all possible three-segment model, all of which rejected the null hypothesis at the 1% level. The IIA property holds (for the full results see the Appendix).

More often the not, the empirical literature on labour market segmentation in developing countries pool together self-employed workers with informal sector wage earners, in order to create the informal sector category (Telles, 1993; Marcouiller et al., 1997; Saavedra and Chong, 1999). In other cases they ignore formal and informal sector differences, and simply focus their analysis on differences among wage earners and entrepreneurs (Blau, 1985; Cohen and House, 1996). The combination of

Figura 3: Occupational Choice and Education.

Figura 4: Occupational Choice and Assets.

such categories can be dangerous, and are only valid if no fundamental underlying differences among the distinct groups are found. The above mentioned literature has failed to notice an important contribution made by Hill (1983) who noticed that the condition for pooling two states of a multinomial logit lies on the equality of the multinomial coefficient vectors, and can be tested through a a likelihood ratio test [11].

In this paper we test for this. Following the procedure suggested by Cramer and Ridder (1991) sequentially for all different possible combinations of our four categories. We were able to reject the possibility pooling any of our categories together at the 1% level for all cases (for the full results see the Appendix).

By analyzing how a set of observable characteristics is distributed across labour market segments, we are able to infer something about the extent of unobserved factors in the search and hiring process. This information is now used to form a set of synthetic regressors that is, in turn, introduced into the earning equations in order to obtain consistent estimates of the determinants earnings(for details, see Model 10).

One extremely important factor when estimating any two-stage model is the identification of exclusion restrictions; in other words, variables which affect the occupation decision, but not affect the earnings equation. In our model, we relied on the existing literature on entrepreneurship and self-employment in order to choose, a set of regressor that fit this quality; in particular, house ownership and partner occupation. The first variable is a proxy of the physical assets of the household, an often important factor when deciding to became an entrepreneur. The second variable, partner occupation, is also often discussed in the literature also as a relevant factor on the occupation decision[12].

4.2.2 Estimating the earnings equation

The second stage of this economic procedure involves estimating earnings equations of each segment including the additional synthetic regressors (Model 10). The chosen earning equation is a variant of Mincer's standard human capital equations, and included controls for education, experience and experience squared, marital status, migration, work journey, sector, household position (head of the household), and neighborhood average income and distance to downtown Rio.

As usual, the dependent variable in this analysis is the natural logarithm of monthly earnings. Each segment's earnings equation also include a separate selection term for each categorical latent expression. Hence, in our four-sector model there are four selection terms, derived from the predicted values of Equation (9), $(\sigma_s\tilde{\rho}_1)$, $(\sigma_s\tilde{\rho}_2)$, $(\sigma_s\tilde{\rho}_3)$, and $(\sigma_s\tilde{\rho}_4)$, where s represents the chosen occupation.

The estimated coefficients are reported in Table 9 for males, and Table 10 for females.

Under the elaborate method, an F-test was conducted to test the hypothesis that all the $(\sigma_s\tilde{\rho}_j)$s are zero. In each segment, the rejection of the null hypothesis[13] indicates the presence of unobservables on the occupational choice process.

For the males (Table 9), we were able to reject the null hypothesis that the $(\sigma_s\tilde{\rho}_j)$s are zero across all four occupations, although the effects of unobservable seems to be stronger for male wage earners than for the entrepreneurs.

For the females (Table 10), we were able to reject the null hypothesis that the $(\sigma_s\tilde{\rho}_j)$s are zero only for the female wage earners.

Overall, these results seems to suggest that the unobservables effects seems to be stronger for the wage earners than for entrepreneurs.

[11] This test was latter reintroduced by Cramer and Ridder (1991)

[12] Although both variables were statistically significant of the first stage regression neither one of them were statistically significant on the second stage.

[13] Note that this is an imprecise test because it ignores the variance from the first stage estimates

The comparison of Tables 9 and 10 shows that females still present a higher return to education than males. Nevertheless, returns to education, for both males and females, are slightly smaller when we correct for occupation selectivity (comparison of Table 6 with 9, and Table 7 with 10). Having said that, it is important to notice that in the case of females wage earners, returns to experience became much higher as we correct for the selection bias.

Tabela 9: **Selected Coefficients Sector Wage Equations Male Workers : BFG**

	wageform	wageinfo	entform	entinfor
	(1)	(2)	(3)	(4)
Education (Yrs of Schooling)	.032	.031	.017	-.025
	(.005)***	(.006)***	(.007)*	(.02)
Experience	-.003	-.008	-.043	-.031
	(.008)	(.009)	(.017)**	(.023)
Experience Squared	.00003	.0001	.0006	.0003
	(.0001)	(.0001)	(.0002)*	(.0004)
Marital status (married)	.043	.091	.017	.165
	(.051)	(.063)	(.085)	(.169)
Migrant	-.05	-.021	-.084	.035
	(.03)	(.034)	(.053)	(.108)
Head of household	-.036	-.088	-.039	.132
	(.046)	(.059)	(.092)	(.155)
Neighborhood Average Income	.00007	.0001	.0001	.0001
	(1.00e-05)***	(.00002)***	(.00003)***	(.00009)
Slum average distance to Downtown Rio	9.16e-07	2.47e-06	5.91e-07	-3.44e-06
	(4.87e-07)	(8.25e-07)**	(1.08e-06)	(2.64e-06)
$(\sigma_s\tilde{\rho}_1)$	1.392	-.322	.998	-2.123
	(.512)**	(.74)	(1.245)	(4.098)
$(\sigma_s\tilde{\rho}_2)$	3.434	.471	2.938	-1.771
	(1.07)**	(.367)	(1.364)*	(3.79)
$(\sigma_s\tilde{\rho}_3)$	1.46	-.555	-.028	.293
	(.859)	(.858)	(.385)	(2.828)
$(\sigma_s\tilde{\rho}_4)$	-.92	-1.801	-1.936	-.946
	(1.462)	(1.448)	(1.675)	(.54)
Obs.	7850	3177	1967	314
$F-$statistic	.213	.31	.229	.309
R^2	111.227	74.668	30.485	6.908
F-test:[a]				
$\lambda_{ij} = 0$	94.03***	48.87***	40.05***	8.16***

Bootstrapped Standard errors in parentheses; Significance levels::10% **:5% ***:1%.*
[a] *Test statistic is approximately* $F_{(4,N-17)}$ *distributed.*
Source:PCBR-Dom

5 Conclusion

This paper provided evidence on the labour market from extremely deprived areas in a developing country, namely the slums of Rio de Janeiro, and examined how it relates to the entrepreneurial activities of the residents of these communities. Some main features are highlighted.

This paper also provides empirical evidence of significant neighborhood effects in the city of Rio de Janeiro, through which the slums closer to more affluent areas of the city had greater earnings, in particular in the case of males.

This paper also applies several tests for the segmentation hypothesis, in particular the geographical and the labour market ones.

Following a well established empirical literature on labour market segmentation on developing countries, this paper shows a substantial earning differential between formal entrepreneurs and formal wage earners, as well as informal entrepreneurs and informal wage earners, with formal sector entrepreneurs earning 15% more and informal entrepreneurs earning 8% less than formal wage earners.

Tabela 10: **Selected Coefficients Sector Wage Equations Female Workers : BFG**

	wageform	wageinfo	entform	entinfor
	(1)	(2)	(3)	(4)
Education (Yrs of Schooling)	.034	.044	-.007	.006
	(.009)***	(.011)***	(.017)	(.026)
Experience	.037	.043	.032	.044
	(.01)***	(.009)***	(.017)	(.023)
Experience Squared	-.0004	-.0005	-.0003	-.0006
	(.0001)**	(.0001)***	(.0002)	(.0003)*
Marital status (married)	.153	.165	.177	.036
	(.09)	(.089)	(.117)	(.229)
Migrant	.049	.059	.046	.196
	(.033)	(.04)	(.053)	(.099)*
Head of household	.127	.146	.132	.123
	(.056)*	(.061)*	(.09)	(.153)
Neighborhood Average Income	.00008	.00004	.00007	.0001
	(.00002)***	(.00004)	(.00005)	(.00009)
Slum average distance to Downtown Rio	-9.88e-08	1.42e-06	1.63e-06	8.09e-06
	(7.66e-07)	(1.58e-06)	(2.19e-06)	(3.46e-06)*
$(\sigma_s\tilde{\rho}_1)$	1.299	2.251	.626	1.209
	(.535)*	(.942)*	(.649)	(1.44)
$(\sigma_s\tilde{\rho}_2)$	.558	.283	.05	-.684
	(.636)	(.246)	(.658)	(1.222)
$(\sigma_s\tilde{\rho}_3)$	2.855	2.333	.72	2.118
	(1.084)**	(1.011)*	(.323)*	(1.683)
$(\sigma_s\tilde{\rho}_4)$	1.383	1.205	.149	.072
	(1.021)	(.842)	(.712)	(.442)
Obs.	3376	958	877	415
$F-$statistic	.19	.317	.281	.41
R^2	41.492	22.884	18.585	16.197
F-test:[a]				
$\lambda_{ij} = 0$	27.79***	9.41***	10.26***	3.93***

Bootstrapped Standard errors in parentheses; Significance levels::10% **:5% ***:1%.*

[a] *Test statistic is approximately* $F_{(4,N-17)}$ *distributed.*

Source:PCBR-Dom

Another important analysis was scrutiny of the process of occupational choice of the population from communities. Two important features, in agreement with the existing literature, emerged from this analysis. Entrepreneurs (both formal and informal) are the only occupational categories in which the probability of entering increases with age, suggesting that this is an extremely important segment for the older population (above 40 years old). Furthermore, it is clear that formal education only has a positive effect on the probability of becoming a formal wage earners, playing a minor role on entrepreneurial choice, and informal wage earnings.

The information from the occupational choice model is implemented on the occupational choice selection terms, which are included in the earnings equation to adjust for unobservables factors that might influence both the occupational choice and the earnings equation. The procedure implemented in this study has never been used in this literature, and has the advantage of including correction term for all occupational possibilities. The coefficients of the selection correction terms were statistically significant, suggesting both the effect of unobservables and the necessity to take them into account on the estimation.

Overall, the present work corroborates other studies of developing countries, showing that returns to education for entrepreneurs is much smaller than for wage earners. However, returns to education found in residents from the slums of Rio are smaller than ones in the developing countries literature. The paper also provides evidence on the residents of the slums of Rio that do not manage to find work outside their communities will, in general, get a substantially lower remuneration for their work. This work also shows that, once we control for selection, the returns for education for formal and informal entrepreneurs is much smaller than for wage earners. Such findings can potentially legitimize focussed interventions aimed to help the labour markets of these communities to better operate, in particular, the entrepreneurial initiatives within this communities.

Acknowledgments

I wish to thank the participants of the EGDI-WIDER conference 'Unlocking Human Potential: Linking the Informal and Formal Sectors' and the UNDP International Poverty Centre seminar series for their helpful comments and suggestions, in particular to those from Fábio Veras, Nanak Kakwani and Benjamin Davis. I also thank the UK Department for Education and Skills for research support under the ORSAS award #2002029001.

Referências

AciolyJr., C. (2001). Reviewing urban revitalization strategies in rio de janeiro: from urban project to urban managment approaches. *Geoforum 32*, 509–520.

Amsberg, J. v. (1999). Rio: Poverty in the municipality. In I. Gill, S. Yusuf, J. v. Amsberg, W. Dillinger, B. Menezes, and M. Neri (Eds.), *Rio de Janeiro: A City Study*, Volume II of *Report No. 19747-BR*, pp. 77–105. Washington D.C.: Word Bank.

Banerjee, A. V. and A. F. Newman (1993). Occupational choice and the process of development. *Journal of Political Economy 101*(2), 274–98.

Becker, G. S. (1964). *Human capital : a theoretical and empirical analysis,with special reference to education*. New York: Columbia University Press for National Bureau of Economic Research.

Becker, G. S. and B. Chiswick (1966). Education and the distribution of earnings. *American Economic Review 56*, 358–359.

Bernhardt, I. (1994). Comparative advantage in self-employment and paid work. *Canadian Journal of Economics 27*(2), 273–89.

Blanchflower, D. G. and A. J. Oswald (1990, February). What makes a young entrepreneur?

Blau, D. M. (1985). Self-employment and self-selection in developing country labor markets. *Southern Economic Journal 52*(2), 351–63.

Borjas, G. J. (1986). The self-employment experience of immigrants. *Journal of Human Resources 21*(4), 485–506.
Borjas, G. J. and S. G. Bronars (1989). Consumer discrimination and self-employment. *Journal of Political Economy 97*(3), 581–605.
Bourguignon, F., M. Fournier, and M. Gurgand (2001, December). Selection bias correction based on the multinomial logit model.
Cain, G. G. (1976). The challenge of segmented labor market theories to orthodox theory: A survey. *Journal of Economic Literature 14*(4), 1215–57.
Calvo, G. A. and S. Wellisz (1980). Technology, entrepreneurs, and firm size. *Quarterly Journal of Economics 95*(4), 663–77.
Cira, D. A. (2002). Urban upgrading in latin america and the caribbean. *En Breve - World Bank* (3).
Cohen, B. and W. J. House (1996). Labor market choices, earnings, and informal networks in khartoum, sudan. *Economic Development and Cultural Change 44*(3), 589–618.
Cramer, J. S. and G. Ridder (1991). Pooling states in the multinomial logit model. *Journal of Econometrics 47*(2-3), 267–72.
De Soto, H. *The mystery of capital : why capitalism triumphs in the West and fails everywhere else.* New York: Basic Books.
De Soto, H. (1989). *The other path : the invisible revolution in the Third World* (1st ed.). New York: Harper & Row.
De Wit, G. and F. van Winden (1989). An empirical analysis of self-employment in the netherlands. *Small Business Economics 1*, 263–272.
Dubin, J. A. and D. L. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica 52*(2), 345–62.
Evans, D. S. and L. S. Leighton (1989). Some empirical aspects of entrepreneurship. *American Economic Review 79*(3), 519–35.
Evans, M. (1989). Immigrant entrepreneurship: Effects of ethnic market size and isolated labour pool. *American Sociological Review 54*(950-962).
Fei, J. C. H. and G. Ranis (1964). *Development of the labor surplus economy: theory and policy.* Homewood, Ill.,: R.D. Irwin.
Ghatak, M. and N. Jiang (2002). A simple model of inequality, occuaptional choice, and development. *Journal of Development Economics 69*, 205–226.
Gill, I., S. Yusuf, J. v. Amsberg, W. Dillinger, B. Menezes, and M. Neri (1999, june). Rio de janeiro: A city study. Technical Report Report No. 19747-BR, World Bank.
Harris, J. R. and M. P. Todaro (1970). Migration, unemployment & development: A two-sector analysis. *American Economic Review 60*(1), 126–42.
Hausman, J. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica 52*(5), 1219–40.
Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–61.
Heckman, J. J. and V. J. Hotz (1986). An investigation of the labor market earnings of panamanian males: Evaluating the sources of inequality. *Journal of Human Resources 21*(4), 507–42.
Hill, M. A. (1983). Female labor force participation in developing and developed countries-consideration of the informal sector. *Review of Economics and Statistics 65*(3), 459–68.
IBGE (2000). Brazilian census. Technical report, Brazilian National Office of Statistics.
ILO (1972). *Employment, incomes and equality : a strategy for increasing productive employment in Kenya : report of an inter-agency team.* Geneva: International Labour Office.
LabGeo (1996a). Mapa digital das favelas do município do rio de janeiro.
LabGeo (1996b). Mapa digital dos bairros do município do rio de janeiro.
Lazear, E. P. (2003). Entrepreneurship. Technical report, IZA Discussion Paper Series no.706.
Le, A. T. (1999). Empirical studies of self-employment. *Journal of Economic Surveys 13*(4), 381–416.
Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica 51*(2), 507–12.
Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *Manchester School*,

139–191.
Lucas, Robert E, J. (1978). On the size distribution of business firms. *Bell Journal of Economics 9*(2), 508–23.
Magnac, T. (1991). Segmented or competitive labor markets. *Econometrica 59*(1), 165–87.
Marcouiller, D., V. Ruiz de Castilla, and C. Woodruff (1997). Formal measures of the informal-sector wage gap in mexico, el salvador, and peru. *Economic Development and Cultural Change 45*(2), 367–92.
Mincer, J. (1958). Investiment in humam capital and personal income distribution. *Journal of Political Economy 56*, 281–302.
Mincer, J. (1974). Schooling, experience and earnings. Technical report, National Bureu of Economics Research.
Parker, S. C. (2004). *The Economics of Self-Employment and Entrepreneurship*. Cambride: Cambrdige University Press.
Ravallion, M. (2002). On the urbanization of poverty. *Journal of Development Economics 68*(2), 435–42.
Rees, H. and A. Shah (1986). An empirical analysis of self-employment in the u.k. *Journal of Applied Econometrics 1*(1), 95–108.
Riley, E., J. Fiori, and R. Ramirez (2001). Favela bairro and a new generation of housing programmes. *Geoforum 32*, 521–531.
Saavedra, J. and A. Chong (1999). Structural reform, institutions and earnings: Evidence from the formal and informal sectors in urban peru. *Journal of Development Studies 35*(4), 95–116.
Schiller, B. R. and P. E. Crewson (1997). Entrepreneurial origins: A longitudinal inquiry. *Economic Inquiry 35*(3), 523–31.
Schmertmann, C. P. (1994). Selectivity bias correction methods in polychotomous sample selection models. *Journal of Econometrics 60*(1-2), 101–32.
Silva, D. B., J. M. Lima, and P. L. Quintslr (1998). A socio-economics survey for job and income development in rio de janeiro slums. In *Proceedings of the Joint IASS/IAOS Conference: Statistics for Economics and Social Development*, pp. 9.
Telles, E. E. (1993). Urban labor market segmentation and income in brazil. *Economic Development and Cultural Change 41*(2), 231–49.
UN-HABITAT (2003). The challenge of slums - global report on human settlements 2003. Technical report.
Van Der Sluis, J., M. Van Praag, and W. Vijverberg (2003). Entrepreneurship selection and performance: A meta-analysis of the impact of education in industrialized countries.