Segmentation in the Brazilian Labor Market*

Fernando Botelho[†] Vladimir Ponczek[‡]

July 24, 2006

Abstract

This paper measures the degree of segmentation in the Brazilian labor market – the wage differential between formal and informal workers – with data from the *Monthly Employment Survey (PME)*, a rotating panel of households in six metropolitan regions. Controlling for observable and unobservable characteristics, we find that workers earn more in the formal sector, which supports the segmentation hypothesis. Contrary to other studies that rely on *two-step procedures* to correct for self-selection bias or *repeated cross-sections* to account for fixed effects, our true panel allows us, under certain conditions, to solve both problems simultaneously, providing a better estimate of the wage differential. We also break down the degree of segmentation by age, gender, position in the household and other socio-economic attributes to identify the groups where this phenomenon is more prevalent. Finally, we investigate the robustness of our findings to the inclusion of self-employed individuals, and also use a two-stage panel probit model to test for a potential weakness of our simple fixed-effects estimator.

Resumo

Este artigo mede o grau de segmentação no mercado de trabalho brasileiro - o diferencial de salário entre trabalhadores formais e informais - utilzando os dado da *Pesquisa Mensal de Emprego* (PME), um painel rotativo de domicílios para as seis maiores regiões metropolitanas do Brasil. Controlando por variáveis observáveis e não observáveis, nós encontramos que trabalhadores recebem mais no setor formal, sugerindo a existência de segmentação. Diferentemente de outros estudos que dependem de procedimentos de dois estágios para corrigir viés de auto-selecão ou *cross-section* repetidas para lidar com efeitos fixo, nosso painel verdadeiro, sob certas condições, permite resolver simultaneamente ambos os problemas, gerando um estimador melhor do diferencial de salários. Além disso, nós calculamos o grau de segmentação condicional em diferentes características como idade, gênero, e posição na família, entre outros atributos sócio-econômicos. Por último, investigamos a robustez dos resultados após a inclusão na amostra de profissionais conta-própria e usamos um modelo probit em dois estágios para testar potenciais problemas no modelo de efeitos fixos simples.

keywords: labor market, segmentation, fixed effect.

palavras-chave: mercado de trabalho, segmentação, efeito fixo.

Área: 12

JEL: J42, J31, C23, C25

*The authors would like to thank Avinash Dixit and Anne Case for guindance and encouragement. Diego Nehab provided unvaluable programming support on the LUA scripting language. Jorge Takeshta was extremely helpful in making the original dataset available remotely. Fabiana de Felício provided us with essential hints about the structure of the PME survey. Fernando Botelho kindly thanks CAPES for financial support. The usual disclaimer applies.

[†]Ph.D candidate at Princeton University, Princeton, NJ and Visiting Fellow at the University of São Paulo, Brazil. email: `fbotelho@Princeton.EDU`

[‡]Ph.D candidate at Princeton University, Princeton, NJ. email: `vponczek@Princeton.EDU`

1 Introduction

A salient feature of the Brazilian labor market is the existence of a significant number of unregistered workers.[1] Formal and informal positions differ in several dimensions. While registered (or formal) workers usually have access to several fringe benefits, individuals employed in the underground economy lack the access to some of them. According to the model proposed by Rosen (1986), in a frictionless labor market, average earnings should be higher in the less 'desirable' informal sector to compensate for the non-pecuniary benefits granted to registered workers. That is, assuming that the value of the benefits is non-negative, direct payments in the informal sector should be at least as high as those in the formal economy to offset the lack of benefits. Otherwise, workers would choose only jobs in the above-ground economy, increasing the supply of labor to formal firms and driving down the wages in the legal economy. The opposite would happen in the informal sector, arbitraging away the wage differential favoring formal jobs. The coexistence of both markets with a wage differentials favoring formal workers points out to barriers to entry into the formal sector or some other friction that prevents workers from moving to high-paying jobs. In this case, we say that the market would be *segmented*. This papers measures the degree of segmentation in the Brazilian labor market by comparing wages of workers as they move across sectors, and finds that the same workers earns significantly less in the informal sector, suggesting that the market is segmented, an empirical finding in need of explanation.

From the point of view of economic efficiency, a segmented market fails to optimally allocate workers to firms. There are two sources of inefficiency. Assume that firms hire workers up to the point where the marginal value of the productivity equals the prevailing wage rate. In this case, there would be workers currently employed in the low-productivity informal sector willing to supply labor to high-productivity firms in the formal sector at a lower cost (wages plus benefits). A segmented market (caused by barriers to entry) prevents this efficient movement of workers across firms in different sectors, maintaining the productivity differential. The second source has to do with the wage/benefits composition in the formal sector. The value attached to the fringe benefits by the workers may be lower than the cost t hat the firms must incur to provide them. Therefore, those workers would be willing to exchange benefits for a higher wage, that would cost less for the firms. Even though segmentation could arise due to asymmetric information between firms and workers in an efficiency-wage framework,[2] in the particular case under study in this paper, namely segmentation across formal and informal *sectors* in Brazil, the usual suspect is the labor legislation. These observations make the measurement of the degree of segmentation an important issue for policy-makers.

The issue at hand – namely the presence of segmentation – is essentially an empirical endeavor, but not a simple one. Due to observable and unobservable characteristics, unconditional comparisons of wage means would be misleading. For instance, workers in the formal sector may be more educated compared to their peers in the informal sector. Therefore, part of the wage differential would be explained by schooling differences. A naïve approach to estimate the wage advantage of the formal workers by simply comparing means would result in an upward-biased estimate of the segmentation degree. One way to get around this problem is to add controls in a regression framework. In the literature, the variables usually added to the regression as controls include gender, experience, and race.[3]

[1]The expressions *informal* and *unregistered* are used exchangeably to refer to individuals employed in the informal sector, without proper registration. The empirical counterpart of this definition is discussed below. It is also important to mention that, in this paper, *informal sector* refers to a feature of the *jobs* and not to the *firm* in which the worker is employed. Even though we do not have reliable information about the firms (employers) in this survey, one can conjecture that most informal workers are indeed employed by informal firms.

[2]See Shapiro and Stiglitz (1984) for a model in which segmentation and involuntary unemployment arise in equilibrium.

[3]Table 7 presents some descriptive statistics for both groups in our dataset.

A trickier issue that one has to tackle in order to identify the degree of segmentation is the endogeneity caused by unobservable characteristics of the individuals, like ability and intrinsic preferences. The presence of unobservable attributes may cause bias in the estimates, since these characteristics may jointly determine the sector in which the individual will eventually work and his/her earnings. For instance, a more risk-averse individual tends to prefer the formal sector in which the probability of dismissal is lower, and in case of lay-off, one is eligible for unemployment insurance. Furthermore, it is possible that risk averse workers are more prone to effort, resulting in higher wages.

In the literature, the endogeneity of the sectoral choice is dealt with by methods that address selection bias (a Heckman two-stage procedure). This procedure consists in estimating a sectoral choice equation in the first stage (by specifying a probit model), and constructing the correction terms to be used as regressors in the main wage regression. To follow this approach it is important to have variables that can be simultaneously included in the first stage and excluded in the main regression.[4] These excluded variables must have two properties. They should be orthogonal to the errors of the main equation and also relevant to determine the sectoral choice in the first stage. In the datasets available in Brazil it is hard to think of a variable satisfying both conditions.

Our approach explores the panel data structure of the monthly employment survey (*PME*) conducted by the Brazilian national statistics agency *IBGE*. Our longitudinal dataset allows us to keep track of the same individual over several periods. Since the Brazilian labor market is characterized by significant labor force turnover, we observe several *intersectoral transitions* over time. Compared to a cross-section structure, our approach allows us to calculate the wage variation during a transition from one sector to the other for the *same* individual. This strategy is consistent even if there are *unobservable* attributes which simultaneously determine sectoral choice and earnings as long as those characteristics are constant over time.[5] The previous studies that use a cross-section structure cannot control for unobservable attributes and consequently will obtain biased estimates when the observable characteristics jointly explain sectoral choice and earnings. Nevertheless, the fixed-effect estimator has a potential flaw.

This paper is organized as follow. Section 2 reviews the literature related to this topic. A brief overview of the Brazilian labor market is presented in section 3. Section 4 talks about our database. Our empirical strategy is presented in section 5. Section 6 presents and discusses the results. Section 7 concludes and points out direction for further research in the topic.

2 Literature Review

The study of market segmentation is not a new theme in the literature, and particular attention has been devoted to the occurence of this phenomenon in the labor markets, in developing as well as developed countries.[6] Our main concern in this paper is the measurement of the wage differentials between similar workers in different sectors, which is one of the central questions addressed by this literature. As pointed out by several authors, significant wage differentials that persist after controlling for individual attributes indicate the existence of barriers to entry in the formal/primary sector.[7] Broadly speaking, there are two views of the informal labor market. One of them considers

[4]Even though this model can be identified without the exclusion of any variable from the main equation due to the non-linearity in the first stage, this procedure is considered unreliable since the correction term tends to be highly correlated with the other covariates, specially in the range in which the inverse Mills ratio is almost linear, inflating the standard errors, and reducing considerably the accuracy of the estimators.

[5]Indeed, we just need the variation of the unobservable attributes not to be jointly correlated with the earnings and sectoral choice variation.

[6]Leontaridi (1998) provides a comprehensive survey of this literature, with some reference to the econometric issues involved in the measurement of segmentation.

[7]Another important branch of this literature investigates how one's attributes, like education and experience, are valued differently in the two sectors. In this respect, there is some evidence that the returns to education are lower in the informal sector. Since we estimate wage differentials in a panel data framework and there are only a few changes

informal work as a transitory phase for entrants in the labor market, and a buffer between formal employment and unemployment, being a transitory state during recessions. Therefore, lacking the ability to cross the barrier into the formal sector, where the good and desirable, but rationed, jobs are, workers take advantage of the flexibility of precarious positions in the underground economy. Due to institutional barriers (labor regulations, minimum wage, and unions), wages of formal worker do not fall to clear the market, and informal workers would prefer a positions held by their counterparts (with similar attributes) elsewhere. Several other authors, including Maloney (2004), point out that the benefits of being in the formal sector are ellusive, since public services offered to formal worker are inefficient. According to this view, informal jobs may be a desirable alternative, providing more flexibility, and allowing both sides of the market to avoid cumbersome and expensive regulations – and taxes. In this scenario, one could not assert that the labor market is segmented, since unregistered workers may prefer their current positions, sorting themselves into the informal sector. In the Brazlian case, even though the jobs in the informal sector may provide benefits invisible to the outside observer, we believe that formal sector positions are more desirable (even after holding observable pecuniary and non-pecuniary rewards constant). In this case, the test of the segmentation hypothesis hinges on the magnitude of wage differentials, with a wage premium in the formal sector supporting the first view.

Several authors have proposed a measure of the degree of segmentation, for Brazil as well as other LDC's.[8] The empirical strategies can be divided into three categories: cross-sectional studies that correct for self-selection bias, repeated cross-sections that group individuals to control for unobservables, and longitudinal methods applied to panel data. Carneiro and Henley (2001) is an example of the first approach, finding some evidence against the segmentation hypothesis. More precisely, the empirical evidence does not reject a model in which some workers choose sector according to, among other things, the *potential* wage in each sector. Therefore, the wage differential could be explained by differences in personal characteristics of separate sets of workers that make one sector more attractive due to a higher potential wage. The correction for the endogeneity of the occupational choice, a multi-stage estimation procedure, is therefore applied. This strategy consists in estimating a propensity to join each sector in the first stage[9] and using the estimated Mills ratios to correct for the selection bias in the earnings equations. This approach strongly relies on the existence of a set of instruments that explain earning only indirectly, through the likelihood of choosing a sector.[10] Moreover, cross-sectional studies are very susceptible to omitted variable bias due to the presence of unnobservable attributes. Magnac (1991) adopts a similar approach to investigate market segmentation in Colombia, failing to reject the null hypothesis that the market is integrated.

Filho *et al.* (2004), using a repeated cross-sections of Brazilian household surveys ($PNAD$), also concludes against the segmentation conjecture. Grouping the individuals according to age (birth cohort), education, and time, the authors try to control for unobservable characteristics. By constructing a *pseudo-panel* and controlling for group fixed effects, the authors find that workers in the formal sector do not earn more than their counterparts in the informal market. Indeed, they find the opposite: informal workers earn much higher wage than similar workers in the above groud economy. This informal wage premium could be a compensation for the foregone benefits in the formal sector. The weakness of the method lies on its limited ability to control for unobservable attributes. With a genuine panel, we can explicitly control for unobservables which are constant over time.[11]

---

in educational attainment within individuals, there is not much we can say about this question without losing the ability to control for unobservable attributes.

[8]Ulyssea (2005) is a comprehensive review of the literature addressing the informal market and the measurement of the degree of segmentation in Brazil.

[9]One can interpret the equation of the first stage as determining the relative benefit of joining each sector, conditional on personal attributes.

[10]As discussed before, in the absence of instruments one is forced to rely on non-linearities in order to identify the main equation, which considerably reduces the precision of the estimator.

[11]Barros *et al.* (1993) apply the same technique and reach similar consclusions.

Closer to the approach adopted here, Maloney (1999) uses data from the *Encuesta Nacional de Empleo Urbano*, which, like the *PME*, is a rotating panel of urban households, to gauge the degree of segmentation in the Mexican labor market. As argued below, the longitudinal structure allows the author to examine how the wage rate changes when the worker moves between formal and informal employment. Contrary to our conclusion, the Mexican labor market seems to be more integrated, with workers who transit to informal salaried work experiencing a pay increase. Therefore, as pointed out in the paper, it is hard to access to what extent the wage premium in the informal sector compesates for the loss of benefits. This conclusion supports the view that the informal paid work may be a viable option for certain groups – namely young, less skilled workers – who want to escape taxes and regulations that plague the formal sector. Pratap and Quintin (2006) conclude that the evidence cannot reject the hypothesis that the Argentinian labor market is integrated. Applying a semi-parametric methods to a panel of Argentinian households, the authors found the wage differential between formal and informal employmet to be not significantly different from zero. Like the Mexican case, one cannot assert whether the absence of a wage premium in the informal sector compensates the absence of benefits of formal sector employment (if any).

3 Overview of the Brazilian Labor Market

Among the salient features of the Brazilian labor market, the high number of unregistered workers is the one that deserves most attention. Approximately 26% of the workers in our sample are unregistered, not including the unregistered self-employed. The stringent labor legislation and the several macroeconomic crisis that affected the country in the last twenty years, with alternating periods of recession and high inflation, are usually blamed for this widespread phenomenon.[12]

Since labor regulations are pervasive and leave little room for direct negotiation between workers and firms, full compliance with the labor codes is expensive and cumbersome. A good indicator of the degree of segmentation is whether the worker has a 'signed' work booklet (*Carteira Nacional de Trabalho*), an identification card issued by the Ministry of Labor. The terms of the contract between the employee and the employer are supposed to be registered in this booklet, and by making the registration (signing the booklet), the employer automatically agrees to comply to the clauses of a standard labor contract, as defined by the Brazilian Labor Code (*Consolidação das Leis do Trabalho*), written in 1943 and slightly modified since then.[13] Among other things, that statute dictates that the worker is entitled to several benefits, like a thirteenth wage to be paid sometime between November and December, a one-month paid vacation, severance payment for unjustified dismissal (which must be communicated to the worker one month in advance), work week of forty four hours, at least fifty percent premium for overtime work, food and transport subsidy, and a four-month paid pregnancy leave of absence for women. Moreover, wages must be at least as high as the minimum wage. There is a statute of limitations that allows workers to sue the former employers for violations up to five years after the contract was terminated.[14]

As we mentioned in the introduction, there are two types of inefficiencies imposed by this legislation. On the one hand, the combination of wages and benefits may not be the most efficient: workers and firms would have an incentive to negotiate another mix of wages and benefits, cheaper for firms to pay and utility-improving for workers. On the other hand, by imposing a minimum wage which can be binding for less productive workers, the legislation may be destroying jobs and creating inefficiency, as argued extensively in the literature about the impact of the minimum wage. If the

[12]Increased trade openness during the 90's may also have contributed to increased informally by increasing the relative importance of the services sector – where informal employment is more prevalent – relative to manufacturing.

[13]The Federal Constitution that went into effect in 1988 changed some aspects of the labor legislation, modifying, among other things, the regulation concerning the duration of the work week and severance payment for unjustified dismissal.

[14]A comprehensive description of the legislation in Brazil is provided in Amadeo *et al.* (2000).

prescriptions of the labor code are inneficient, leaving room for workers and firms to bargain and reach agreements in the shadow of the law, how can we have market segmentation? In other words, why do firms end up creating formal jobs instead of hiring informal workers for a lower wage, bidding up wages in the underground sector?

One conjecture points to how labor disputes are adjudicated in Brazil. As dictated by the Labor Code, a special branch of the judiciary is responsible for cases involving labor disputes. Moreover, contracts that do not satisfy all the provisions of the Labor Code are considered void, and workers can sue their employers for non-compliance with the legislation. The burden of proof falls on the employers, that have to prosent convincing evidence that the terms of the contract were fulfilled. If a certain worker can prove that the labor relationship ever existed – and the standard of proof is low, being sufficient to provide an eye-witness — the firm is supposed to prove that the requirements imposed by the legislation were fulfilled. Otherwise, the employers may be subject to a fine and payment of compensation to the worker. Therefore, violations of the labor law is the employers' responsibility, even if an informal agreement was reached in advance between the parts. Even though spontaneous detection by the authorities is rather unusual, this labor law branch of the judiciary has been very active, adjudicating about one million cases in 2003 in the states analyzed in this paper. This possibility of suing the employer after the relationship has been severe may deter some firms and (potential) employers from resorting to informal arrangement.

4 The Data

The *Pesquisa Mensal de Emprego* (PME), or Monthly Employment Survey, is a monthly rotating panel of dwellers in six major metropolitan areas in Brazil (São Paulo, Rio de Janeiro, Belo Horizonte, Salvador, Porto Alegre e Recife). Together these six metropolitan regions encompass approximately 30% of the population. We used data from 1995 to 2001[15] to avoid the high inflation period before the monetary stabilization plan in 1994 (*Plano Real*).[16] This dataset is compiled by the Brazilian national statistical agency – *Instituto Nacional de Geografia e Estatística (IBGE)*.

The survey investigates schooling, labor force, demographic, and earnings characteristics of each member of the household age 10 and over, for every interviewed household. Approximately 100,000 individuals in 35,000 households are interviewed every month. Households are interviewed once per month for four consecutive months, then there is an eight-month window when they stand by; after this period, the household is interviewed for another four-month period. For instance, suppose that the first interview was conducted in January. The second, third, and fourth interviews will take place in February, March and April of the same year, respectively. From May to December the household will rotate out of the sample. From January to March of the following year the household is interviewed again and, after this spell, the household is permanently excluded from the sample. Since the main purpose of this survey is to measure the unemployment in the major metropolitan areas in the country, individuals are not directly identified, but only their households. Therefore, we match individuals within households over time using gender, month and year of birth, as well as the *IBGE*'s identification codes. The main questions for the purpose of our study are earnings and the hours worked in the month of reference, the legal status (with or without signed work booklet), the sector of activity, and some variables such as age, gender and schooling.

To avoid seasonality issues, we use observations exactly one year apart. Our regressions include only the first and the fifth interviews. Since we lose track of some individuals along the interview period, this strategy maximizes the number of usable observations. The problem of attrition is usually found in longitudinal surveys, and the PME is not an exception. It is challenging to follow individuals

[15] December/1995 to September/2001.

[16] Since our identification strategy relies on intertemporal comparisons, it is also important to deflate the nominal variables. Even though this is less of a problem after 1995 after the fall of the inflation rate, we deflate the data using the deflator proposed by Corseuil and Foguel (2002).

for sixteen months in the urban environment of a developing country where most of the people do not own their dwellings, and move-overs are frequent. This effect could bias our estimates if the individuals who stay in the sample longer have different responses to change in the formal-informal wage gap. For example, since the household – not the individual – is the sampling unit, every time an individual moves out from the original dwelling, (s)/he is excluded from the sample. Moreover, our matching method to track individuals may not be foolproof. That is, due to coding errors, we may be unable to track the individual over time if one of the variables used to match him/her varies across interviews. This second source of attrition is less harmful since it is more likely to be unrelated to the outcome of interest.[17] The first cause of attrition could be more problematic, since the decision to move out is less likely to be orthogonal to both sectoral choice and earnings. In order to gauge the importance of this source of attrition, we compare subsample composed by individuals with eight interviews ('low' attrition) with a subsample formed by individuals who dropped out of the sample after the fifth interview ('high' attrition).[18] Other feature of the survey, which makes it particularly prone to measurement error, is the fact that the information about some individuals is provided by another dweller, usually the head of the household. To investigate whether these observations are a source of bias, we compare the segmentation in a subsample of self-reported individual with the remaining of the sample.

Finally, we consider only employees in the private sector and exclude self-employed workers in the benchmark specification. First, wages in the public sector are dictated by factors other than market forces. Second, even though the so-called self employed constitute a significant - and growing - fraction of the labor market, this category is extremely heterogenous, including individuals as distinct as street vendors and doctors. Furthermore, self-employed workers are supposed to satisfy different criteria and regulations to join the formal sector.[19] As a robustness test, we run similar regressions including the self-employed in the sample.

5 Empirical Strategy and Identification

As we mentioned in section 3, the issue of identifying segmentation hinges on the wage differential between formal and informal workers, with a positive difference in favor of formal workers suggesting the presence of barriers to entry. Therefore, the main goal is to devise a strategy to measure that magnitude appropriately and to identify groups of individuals where the phenomenon is more prevalent. Most studies in the literature use a cross-section framework, and rely on variations *across* individuals to estimate wage differentials. Instead, by using a panel structure, we directly control for the total individual fixed effect, using the wage variation of the *same individual* to measure the wage gap. First, we estimate the benchmark equation

$$\log(w_{i,t}) = \alpha_0 + \theta\, b_{i,t} + \beta'\, X_{i,t} + e_{i,t} \tag{1}$$

where $X_{i,t}$ is a vector of observable covariates for individual i in period t. The vector $X_{i,t}$ will typically include variables like the sector of employment (to control for wage inequalities among industries),

[17]The results reported here are based on a filter that considers V101, V102 and V103 (*IBGE*'s id codes), V202 (gender), V206 (day of birth), V236 (month of birth), and V246 (year of birth), and schooling (V208, V209, and V210). Out of the 156,233 individuals who had a first interview, 117,282 were located sucessufully one year after, resulting in an attrition rate of 38.5%. Nevertheless we tried several different combinations of variables to match individuals across time, including day, month and year of birth, gender, and position in the household. The results obtained with the various specifications do not significantly differ from each other. The results are available upon request.

[18]As we argue later, the conclusions of this exercise may be misleading, since most of the attrition occurs during the window – between fourth and fifth interviews, and individuals that drop out of the sample between the fourth and fifth interviews may be intrinsically different from the one that do it later.

[19]Self-employed workers are required by law to pay income taxes and social security contribution. The other fringe benefits, as one might expected, are not mandatory.

and a *dummy* variable indicating the year (to eliminate any aggregate trend in the wage rate). Notice that the fixed effect framework does not allow us to include other variables that are constant over time, like gender and race. Also, since the universe of analysis includes only individuals over 16 (above school age), education does not change over time for most of the individuals. Therefore, it is hard to include a measure of schooling as a covariate. The same is true for other variables like position in the household.

It is also possible to assess how the degree of segmetation changes in the several subgroups that make up the sample. For instance, we woud like to know whether the pattern is heterogeneous for different metropolitan regions or industries of occupation. Our regression strategy to carry out this analysis is based on the equation

$$\log(w_{i,t}) = \alpha_0 + b_{i,t}\,(\theta + \phi'\,V_i) + \beta'\,X_{i,t} + e_{i,t}, \tag{2}$$

where V_i is a vector of time-invariant attributes. The coefficient vector of coefficients ϕ measures how the degree of segmentation depends on V_i. Therefore, we can investigate whether the degree of segmentation is more pervasive is certain subgroups of the population.

Back to the fixed effect approach, we assume that the error term in (1) can be further decomposed into

$$e_{i,t} = \mu_i + \epsilon_{i,t} \tag{3}$$

with μ_i being the individual fixed effect (constant over time), and $\epsilon_{i,t}$ is a random error term. The crucial identifying assumption is that $\epsilon_{i,t}$ is uncorrelated with $b_{i,t}$ or $X_{i,t}$. That is, any variable that is not constant over time and is correlated with earnings cannot be relevant for the sectoral choice. More specifically, the consistency of fixed-effect estimator hinges on the assumption that conditional on the covariates and any fixed characteristic of the individuals, the selection process is uncorrelated with the earnings. Assume the following sectoral choice process:

$$b_{it} = 1_{(\gamma Z_{it} + u_{it} \geq 0)}, \tag{4}$$

where Z_{it} is a set of variables explaining sectoral choice. A sufficient condition for our identification assumption to be valid is

$$E(u_{it}\,e_{it}) = 0. \tag{5}$$

However, this condition is too strong, and a much weaker assumption is sufficient for consistency in our case. Let

$$u_{it} = \alpha_i + \upsilon_{it}, \tag{6}$$

where α_i is a fixed effect that captures individual characteristics that influence the sectoral choice. Even in case α_i and μ_i are correlated, as long as υ_{it} and ϵ_{it} are not, $\hat{\theta}$ will be a consistent estimator of the degree of segmentation. This strategy is consistent even if there are *unobservable* attributes which simultaneously determine sectoral choice and earnings as long as those characteristics are constant over time.[20] The previous works which use a cross-section structure cannot control for unobservable attributes, and consequently will obtain biased estimates when the observable characteristics jointly explain sectoral choice and earnings. Moreover, our approach does not require assumptions about the distribution of $\epsilon_{i,t}$, since we are directly conditioning in the individual fixed effect. This is another advantage of the panel data structured relative to two-step estimators that use cross-section data and require a known joint distribution of the errors in both stages.

Nevertheless, the fixed-effect estimator has a potential flaw. Whenever the relevant unobservable attributes changes over time and their variation is correlated with the earnings and sectoral choice variations, the fixed-effect estimator is inconsistent. For example, if workers can forecast their relative prospects in both sectors and make the sectoral choice decision based on those expectations (which

[20] Indeed, we just need the variation of the unobservable attributes not to be jointly correlated with the earnings and sectoral choice variation.

are not observed), then the transition from one sector to the other may also be endogenous.[21] One way to remedy this flaw is to use the method described in Botelho and Ponczek (2006). Briefly, this approach consists in a two-step procedure: in the first step, the endogenous binary variable is modeled as a random-effect probit, being explained by an exogenous instrument. Next, the second stage includes also correction terms constructed to account for the endogenous switching between sectors. We can measures the strength of the instrument in the first stage: a good instrument is relevant to explain sectoral choice, but does not affect earnings directly. In the absence of the endogenous switching problem, both estimators are consistent, however the pure fixed-effect model is more efficient. A straighforward way to check for the presence of endogenous switching is to test the significance of the correction terms in the second stage. Under the null hypothesis that the binary variable is exogenous, the coeficient of the correction term is assymptotically distributed as a zero mean normal and its significance can be assessed by the usual t-test.

In summary, the fixed effect estimator will build consistent estimates of the degree of segmentation even in the presence of potential selection bias when the source of bias is a correlation between formal/informal sector and unobserved factors which made differ across individuals but is constant over time for each individual. If selection into the sector takes this form, the simple fixed-effect estimator is a straightforward way to deal with selection bias.

6 Results

This section presents the main findings and briefly discusses each one of them. Our main purpose is the measurement of the average segmentation degree in the population. We also try to measure the degree of segmentation of different subgroups determined by social-economic and demographic characteristics. Moreover, we exploit various different specifications as tests of robustness. Sometimes they uncover interesting patterns, which are worthy mentioning.

Table 7 presents some simple descriptive statistics. We based the calculations in the first interview of each individual in the sample. In our sample, around 26% of the workers are informal. This number is lower than the national average (around 45%) once it only includes the six largest metropolitan areas of the country, where informalilty is less common. 60% of our sample are males, who work less in the informal sector than women. The construction industry presents the highest proportion of informality. On the other end, the manufacturing industry displays the smallest degree of informality. The disparity in the degree of informality by industry could be associated with differences in type of the labor demand in each industry. For instance, it is possible that the manufacturing industry demands high-specialized workers, which might be more organized and unionized compared to the low-skilled workers in the construction industry. Over the years under study, informality has slightly increased (from 26% in 1995 to 27.2% in 2000) with only small fluctuations every year, which could be capturing a regular activity level variation over the period. Recife/PE, in the Northwest of the country, is the metropolitan area with the highest degree of informality (32%), followed by Belo Horizonte/MG, Salvador/BA, São Paulo/SP, Rio de Janeiro/RJ, and finally Porto Alegre/RS, in the southmost part of the country with only 20% of informal workers. Once again, those regional difference might be associated with difference in the workers characteristics (like union participation) of the workers in each one of the cities. Formal workers have a much higher hourly wage than their informal counterparts. By the same token, the two groups also differ in observable attributes, with formal workers being older (more experienced) and more educated, and working more hours.

To what extent can these differences explain the gap in earnings between sectors? By adding controls in an OLS regression, one can trivially isolated the fraction of the variation in earning that

[21] Ideally, we would like to observe the same individual's earnings, at the time,in both sectors. In this hypothetical situation, the mean wage differential would be a clear and immediate indicator of the degree of segmentation in this labor market. However this strategy is not feasible.

is not explained by the observables. A more subtle issue has to do with non-observable attributes, like talent, motivation, and ability that may be related to sectoral choice and earnings. In this scenario the ordinary regression strategy may fail to obtain a consistent estimator of the degree of segmentation, whereas a fixed-effect estimator may, under certain conditions, account for both issues – self-selection and unobservable variables. Indeed, the random effects model can only produce consistent estimator if the individual effects are not correlated with the other regressors. Table 7 presents the results of a Hausman test, which rejects this hypothesis, concluding that the individual effects are indeed correlated with the regressors and rendering the random effect model inconsistent. Therefore, we adopt a fixed-effect model in the regressions below. In both specifications, we included years and industry *dummy* variables to capture the trend pattern and also to isolate effects caused by changes in the industry from the degree of segmentation. We can see that this is figure is around 7.8% for the period investigated.

Our first set of robustness-check regressions breaks down the transitions into two types – from formal to informal employment, and vice-versa – and investigates the earnings gains or losses associated with the transition. For formal to informal job transitions, individuals that stay in their formal sector jobs are used as a reference group, to account for overall wage gains. For the other type of transition, informal sector stayers are used as controls. The results are presented in table 4 (column 1 for formal to informal transitions). From formal to informal, there appears to be a wage premium of about 7.2% in the formal sector, and 8.4% in the other direction. However, we could not reject the hypothesis that both coefficient are the equal, implying a symmetric effect in the transition from one sector to the other: the gain from transiting from an informal job to a formal one is comparable to the loss one incurs by going the other way.

Table 6 displays three other specifications. In column 1, we included several interaction variables to capture differences in the degrees of segmentation for various subgroups of the population.

Is the segmentation higher among men or women? By looking at the column (1) of the table, one can see that the degree of segmentation is lower among women (-3.6%). Broadly speaking, when transiting from informal to formal employment, earnings increase, on average, 3.6% more for men than for women. This result seems to contradict the intuition that inflexible labor laws are the main cause of the segmentation in the Brazilian labor market, since women in the formal sector have a 4-month paid pregnancy leave of absence as an additional benefit. We wondered whether this seemly odd result could have been driven by differences in the position women usually has within the family in Brazil. Most households with both a man and a women have the man as head of the family and maybe the result of the gender coefficient might be capturing that characteristic. However, this does not seem to be the case. It can been seen at table 9, that the segmentation among male heads is actually significantly smaller (3.9%) than for other males, which does not occur with head females. Moreover, we observe that the segmentation of a spouse female is also smaller than the other females (-5.7%). If one could sort the groups by the degree of segmentation, (s)/he would have that Male (*non* spouse/head $>$ Female (*non* spouse/head $\geq$ male head $>$ female spouse.[22]

Has the degree of segmentation changed during the period of analysis? To investigate this issue we construct a variable which is the interaction between the work status and a *dummy* indicating the year in which the *first* interview took place. By including this variable in the regression, we are allowing the wage differential to differ according to the year of the first interview. Since the base year is 1995, the individuals who had the first interview in 1996 experience a gain of approximately 7% higher than in previous year. For workers who made the transition between 1997 and 1998, this differential rises by 5 percentage points compared to 1995, and so on. 1997 and 2000 present a higher level of segmentation compared to the base year.

[22] We also checked if those results were capturing differences in the number of members in the household headed by a man or a woman by including a set of interaction terms of sectoral choice and *dummy* variables indicating the number of members in the family. The results were very similar to those presented in table 9 and are available upon request.

The bottom of the first part of table 6 presents the degree of segmentation in the metropolitan regions where the survey is undertaken. The reference region in the Greater Recife, which displays the highest level of segmentation. The metropolitan regions of Rio de Janeiro and Porto Alegre seem to be less segmented than the other areas. We could not reject the null hypothesis that Recife, Belo Horizonte, Salvador and São Paulo have the same degree of segmentation.

As mentioned in the introduction, formal sector workers are entitled to receive an extra salary at the end of the year. To prevent our conclusions from being contaminated by the effect of this thirteenth salary, we run the regression with a *dummy* indicating the months of December and January[23]. As can be seen from column three in table 1, those two months display a higher degree of segmentation than the other months (3% more), indicating that over the extra salary at the end of the year indeed account for a fraction of the wag gap between formal and informal workers.

In all regressions, we are controlling for wage differentials between industries to disentangle the industry effect from the formal-informal effect.[24] Nevertheless, we would like also to test if there are disparities in the degree of segmentation for different industries. We perform this analysis by interacting an indicator of industry in the first interview with labor status, for the individuals that were employed in the same industry in both periods. Manufacturing is the reference group. Construction and services present a lower degree of segmentation (around 7% and 5.4%, respectively) and we could barely reject the hypothesis that commerce has the same degree of segmentation as the manufacturing.

Since the worker is not necessarily the person interviewed in the household, we wondered whether measurement errors could be an issue in our study. It is very likely that other members of the families do not know accurately the relevant information, such as earnings, sectoral choice, and other demographics about another family member in the sample. In order to access to what extent information given by other members is biasing our results, we introduced an interaction term with a *dummy* variable that specifies whether the worker him/herself was interviewed. It can be seen in column 2 of table 6 that the interaction term is *not* significant suggesting that information given by other familys member is not misleading our results.

Is attrition a serious problem? As mentioned in section 4, we lose track of 38.5% of our sample from the first to the fifth interview. Therefore, our results capture only the wage premium of those who stay in the sample. As discussed before, this phenomenon could potentially bias the results if there is any characteristic associate to those who stay correlated with the sectoral choice and earnings. In order to investigate if attrition could jeopardize our results, we introduce another interaction term by creating a *dummy* variable that indicates whether the individual has not left the sample after the fifth interview, i.e., if (s)he completed all 8 rounds of interviews. Column 3 of table 6 shows the results of the regression, with the coefficient of the interaction term being non-significant. This indicates that at least attrition after the fifth interview is not correlated with the degree of segmentation. Of course, to accept the claim that attrition is not biasing our results one should believe that the moving decisions made between the first and fifth interviews are the same (at least correlated with the same variables) of those made between the fifth and the eighth interview. However, since those who left the sample only after the fifth interview lived in the same address for at least 12 months[25], this type of dropouts could be very different from those who left before the fifth interview. For instance, since homeowners move much less frequently than tenants, it is more likely that we have more homeowners in the second type of dropouts than in the first one.

Next, we turn to the degree of segmentation over the life-cycle. By adding the interaction between labor status and age we can analyze how the segmentation changes as the individual ages. The

[23]This end-of-year bonus is usually paid in December, and acounted for in January. The results are displayed in the top of the second part of table 6

[24]Even though inter-industry transition are not so usual, the wage differentials paid to workers in different industries may be very significant, jeopardizing our conclusions.

[25]Remember that the fifth interview occurs one year after the first one, subsequent to the 8 months window

interactions between labor status and age bracket *dummies* capture non-linearities in the relationship. Is the transition to informality more painful for younger workers as compared to mature adults? Our result shows that segmentation is actually more painful for young workers. One could conjecture that a policy to make the access of young workers to the labor market should contemplate the question of segmentation, its causes and consequences. To account for other types of non-linearity, we divide the workers in three age brackets (less than 30, 30-49, and more than 49) according to the age in the first interview. Segmentation falls with age, with the difference being statistically significant. We also created three categories to measure schooling achievement. The reference group (*schooling*=0) represents individuals with incomplete fundamental education (including the illeterate). The second group comprises individuals with at least complete fundamental education but no college attendance. Finally, the third group includes all workers with at least some college level education (including college dropouts). By interacting this set of *dummies* with the age groups, we can observe how the life-cycle paterns differ for the various schooling level, and vice-versa. For instance, the baseline group (young individuals with little schooling) is 10 percentage points more segmented than the oldest and more educated group. Overall, segmentation decreases as the individual becomes older and more educted.

One interesting question is to what extent the degree of segmentation is uniform over the entire distribution of the real wages. By creating four separate groups of workers defined by the wage quartile,[26] we can have a better understanding of the pattern of segmentation in different intervals of the wage distribution. To test the significance of the segmentation degree difference between contiguous quartiles, we ran a nested regression (benchmark) including the interactions between sectoral choice and indicators of the wage quartile. An interesting pattern emerges from this table. The segmentation is much larger for low-wage workers, and is *negative* in the top quartile. The results in table 10 indicate that the wage premium for low wage workers is around 15% in favor of the formal sector. The second group has a lower – but still significant – degree of segmentation (7.7%). The third group display no significant difference between wages in the different sectors. Finally, in the top of the distribution one finds evidence supporting that the segmentation is non-existent, with the wage differential being in favor of informal workers, which is predicted by the classical theory. The bottom panel depicts the results of F-tests, which allow us to statistically confirm the difference between adjacent quartiles. More segmentation at the bottom of the wage scale may be consequence of a binding minimum wage constraint for low-productivity workers. On the other hand, the top of the distribution is non-segmented. This result was expected, since the minimum wage is not binding and reputational concerns become more relevant. An informal worker may be more reluctant to seek compensation from the employer through the judiciary system once the informal contract is severed, since the market is more concentrated and the flow of information among employers about an employee's reliability is more intense. Therefore, the cost of resort to courts (in terms of future employment prospectus) is higher. Those results should be taken with caution, since low-wage workers also tend to be younger and less educated. To control for such heterogeneity we run the same regression including interactions of sectoral choice and the possible intervening variables (the same used in table 6). The results, which are available upon request, display the same pattern as before, with segmentation decreasing with wage.

As discussed in section 5, the fixed effect estimator would be biased if the decision to seek a position in the other sector is correlated with the wage/earning variation. In ths case, we have to correct for the bias by finding an instrument: a variable that explains the sectoral choice but is not correlated with the error in the earnings regression. That is, we need a variable that explains earnings only indirectly, through its impact on sectoral choice. In the first stage, we ran a probit model with random effects, calculating the estimated probability of being in one of the formal sector, as a function of the variable *frac*. This variable is construct as the number on children under 10 over the number of members in the household. It can be argued that families with a high number of

[26]The individuals were ranked according to the average wage in the two observed periods.

children will attach a higher value to the benefits granted to formal workers. Therefore, its members will pursue a position in the formal sector more intensively. Indeed we can see at the bottom of table 11 that the variable is significantly positive, *i.e.*, workers who live in a household with more children are more likely to be in the formal sector. The probabilities calculated in the first stage are used to construct the correction terms, which will be included in the main regression. Now, by running a fixed effect regression, including the correction terms, we can observe that the level of segmentation is still significant. Actually, the estimated level of segmentation is even higher (11.5% versus 7.8% in the pure fixed effect estimation). Nevertheless, the correction term is not significant, which suggests that the pure fixed effect strategy is consistent, and, since the two-step procedure is less efficient, one should prefer the more parcimonious model.

Finally, including self-employed workers in our sample does not change much the main conclusions of our study. Column 3 of table 11 shows the results of the benchmark regressions after adding in a variable that identifies if the worker is self-employed or not. We can see that there is a wage premium of 9.5% after changing from the informal sector to a self-employed position. The degree of segmentation between formal and informal barely changes. Comparing the segmentation found in this case with the one found in the benchmark regression without the self-employed workers (column 1), the wage premium slightly increases from 7.8% to 8.3%.

7 Conclusion and Directions for further Research

In this paper we have measured the degree of segmentation in the Brazilian labor market using a genuine panel of individuals, which allowed us to control for unobservable individual attributes that are constant over time. We find the average wage differential between formal and informal workers to be 7.8%, suggesting that the Brazilian labor market is indeed segmented. Moreover, we observe that some individual characteristics have different impacts on the level of segmentation. For instance, females, head of households, older, and more educated workers experience smaller (larger) wage losses (gains) when transferring from formal (informal) to informal (formal) employment. The segmentation seems to be more pervasive among young, low-educated men, who are not the main bread earners of their respective families. Also, the wage differential is much higher in the manufacturing industry and smaller in services. Finally, segmentation seems to be deeper to those in the bottom of the wage distribution, mild to those in the middle and inexistent in the top.

To the best of our knowledge, the strategy used in this paper has never been applied to study this phenomenon in Brazil, and we claimed that our method requires weaker assumptions for consistency than the other approaches used so far to measure segmentation. Nevertheless, as explained in the empirical strategy section, our results rely on the assumption that the sectoral choice is not correlated with wages through the non-fixed part of the error terms. In order to relax that assumption, we applied a two-step estimation method, which estimates a random effect probit in the first method to construct correction terms for the second stage. The second stage still uses a fixed-effect approach and delivers consistent estimator even in the presence of correlation between wage and sectoral choice through non-fixed elements of the error terms. Another big advantage of this approach is that it creates a very straightforward way to test if the simple fixed-effect model utilized in this paper is consistent, i.e., if the only source of endogeneity of sectoral choice comes from individual fixed characteristics. Nevertheless, we can not conclude that the two-step procedure generates better estimates than the pure fixed-effect model, once we the correction term is found to be non-significant.

References

Amadeo, E., Gill, I., and Neri, M. (2000) "Brazil: The Pressure Points in Labor Legislation." Ensaios Econômicos (Fundação Getúlio Vargas, Rio de Janeiro), 395.

Barros, R., Mello, R., and Pero, V. (1993) "Informal Labor Contracts: A Solution or a Problem?" IPEA - Texto para Discussão 291.

Botelho, F. B. and Ponczek, V. P. (2006) "A Note on Endogenous Switching with Panel Data."

Carneiro, F. G. and Henley, A. (2001) "Modelling Formal vs. Informal Employment and Earnings: Micro-econometric Evidence for Brazil." U of Wales at Aberystwyth Management & Business Working Paper No. 2001-15.

Corseuil, C. H. and Foguel, M. N. (2002) "Uma sugestão de deflatores para rendas obtidas a partir de algumas pesquisas domiciliares do IBGE." Discussion Paper 897, Instituto de Pesquisa Econômicas Aplicadas.

Filho, N. A. M., Mendes, M., and oes de Almeida, E. S. (2004) "O Diferencial de Salários Formal-Informal no Brasil: Segmentação ou Viés de Seleção?" *Revista Brasileira de Economia* **58**(2): pp. 235–248.

Leontaridi, M. (1998) "Segmented Labour Markets: Theory and Evidence." *Journal of Economic Surveys* **12**(1): pp. 103–109.

Magnac, T. (1991) "Segmented or Competitive Labor Markets?" *Econometrica* **59**(1): pp. 165–87.

Maloney, W. F. (1999) "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico." *World Bank Economic Review* **13**(2): pp. 275–302.

Maloney, W. F. (2004) "Informality Revisited." *World Development* **32**(7): pp. 1159–1178.

Pratap, S. and Quintin, E. (2006) "Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach." *European Economic Reveiw* (forthcoming).

Rosen, S. (1986) *The Theory of Equalizing Differences*, vol. 1, chap. 6, pp. 641–692. Elsevier Science.

Shapiro, C. and Stiglitz, J. E. (1984) "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* **74**(3): pp. 433–444.

Ulyssea, G. (2005) "Informalidade no Mercado de Trabalho Brasileiro: Uma Resenha da Literatura." Tech. Rep. 1070, Instituo de Pesquisas Econômicas Aplicadas - IPEA.

Appendix – Tables

Table 1: Descriptive Statistics

Variable	**Count**		**Percentage**	
	Informal	**Formal**	**Informal**	**Formal**
sample size	40,398	115,835	25.9%	74.1%
			by gender	
male	22,328	71,636	23.8%	76.2%
female	18,070	44,199	29.0%	71.0%
			by head of household	
yes	18,407	64,120	22.3%	77.7%
no	21,991	51,715	29.8%	70.2%
			by industry	
manufacturing	5,546	29,322	15.9%	84.1%
construction	3,710	5,556	40.0%	60.0%
commerce	4,948	17,248	22.3%	77.7%
services	25,273	63,096	28.6%	71.4%
			by year/wave	
1995/96	11,646	34,759	25.1%	74.9%
1997/98	13,853	41,406	25.1%	74.9%
1999/00	14,899	39,670	27.3%	72.7%
			by state	
Recife/PE	5,609	11,842	32,1%	67.9%
Salvador/BA	5,222	14,075	27.1%	72.9%
Belo Horizonte/BH	10,063	23,578	29.9%	70.1%
Rio de Janeiro/RJ	7,628	23,966	24.1%	74.9%
São Paulo/SP	9,217	26,017	26.2%	73.8%
Porto Alegre/RS	5,342	20,218	20.9%	79.1%
	Mean		**Difference**	**Standard Error**
	Informal	**Formal**	(Formal-Informal)	
wage	1.866	2.63	0.770	0.0237
hours worked	41.523	41.788	0.265	0.0656
age	33.502	34.734	1.233	0.0618
years of schooling	6.887	8.067	1.179	0.0231

Note: Except for the variable *wage*, statistics were calculated with the first interview only. The *wage* variable was constructed as the average of the wage in the 1th and 5th interviews.

Table 2: Wage within quartiles

Quartile	mean	std. dev.	min	max	% of Informal
1st quartile	0.627	0.161	0.026	0.882	41.42%
2nd quartile	1.127	0.150	0.882	1.408	23.08%
3rd quartile	1.862	0.312	1.408	2.525	18.79%
4th quartile	6.245	5.814	2.525	265.95	17.35%

Table 3: Hausman Test

	(b) fixed-effects	(B) random-effects	(b)-(B) Difference	$(V_b - V_B)^2$ S.E.
$formal$	0.0780	0.2152	-0.1372	0.0030

fixed-effects (b): consistent under H_0 and H_a

random-effects (B): inconsistent under H_a, efficient under H_0

H_0: difference in coefficients is not systematic.

$\chi^2_{(11)} = (b - B)'[V_b - V_B]^{-1}(b - B) = 3,319.07$

p-value=0.0000

Note: log of real wages is the dependent variable. Regressions include *year*, *sector*, and *individual fixed effects*. Students and individuals with schooling transition were excluded.

Table 4: Direction of transition

Variable	**Coefficient** (Std. Error)
informal to formal ($\hat{\beta}_1$)	0.07235 (0.00633)
formal to informal ($\hat{\beta}_2$)	0.08487 (0.00697)
$H_0 : \beta_1 - \beta_2 = 0$	
$\hat{\beta}_1 - \hat{\beta}_2$	-0.01252 (0.00974)

Note: log of real wages is the dependent variable. Regressions include *year*, *sector*, and *individual fixed effects*. Students and individuals with schooling transition were excluded.

Table 5: Wage within quartiles

Quartile	mean	std. dev.	min	max	% of Informal
1st quartile	0.627	0.161	0.026	0.882	41.42%
2nd quartile	1.127	0.150	0.882	1.408	23.08%
3rd quartile	1.862	0.312	1.408	2.525	18.79%
4th quartile	6.245	5.814	2.525	265.95	17.35%

Table 6: Robustness check

Variable	**Coefficient** (Std. Error)		
	(1)	(2)	(3)
formal	0.05726 (0.03536)	0.05726 (0.03536)	0.04288 (0.03709)
formal × self-reported	–	-0.00079 (0.01591)	–
formal × complete	–	–	0.01777 (0.01386)
formal × male	0.03692 (0.01576)	0.03690 (0.01577)	0.03687 (0.01576)
formal × head	0.01441 (0.01547)	0.01452 (0.01561)	0.01456 (0.01547)
formal × male × head	-0.05341 (0.01930)	-0.05348 (0.01934)	-0.05353 (0.01930)
formal × spouse	-0.03727 (0.01526)	-0.03721 (0.01531)	-0.03741 (0.01526)
formal × male × spouse	-0.01936 (0.05014)	-0.01943 (0.05016)	-0.01924 (0.05014)
formal × 1996	0.06958 (0.02954)	0.06958 (0.02954)	0.06917 (0.02954)
formal × 1997	0.11890 (0.03109)	0.11891 (0.03109)	0.11847 (0.03109)
formal × 1998	0.08548 (0.02964)	0.08549 (0.02964)	0.08483 (0.02964)
formal × 1999	0.09918 (0.03079)	0.09919 (0.03079)	0.09862 (0.03079)
formal × 2000	0.10478 (0.02953)	0.10478 (0.02953)	0.10610 (0.02955)
formal × Salvador/BA	-0.02021 (0.01921)	-0.02018 (0.01922)	-0.02063 (0.01922)
formal × Belo Horizonte/MG	-0.01818 (0.01719)	-0.01814 (0.01721)	-0.01882 (0.01720)
formal × Rio de Janeiro/RJ	-0.04120 (0.01778)	-0.04117 (0.01779)	-0.04174 (0.01778)
formal × Sâo Paulo/SP	-0.01070 (0.01705)	-0.01068 (0.01705)	-0.01107 (0.01705)
formal × Porto Alegre/RS	-0.05034 (0.01819)	-0.05029 (0.01823)	-0.05111 (0.01820)

Table 6: Robustness check (contd.)

Variable	**Coefficient** (Std. Error) (1)	(2)	(3)
formal × Dec/Jan	0.02961	0.02961	0.02905
	(0.01325)	(0.01325)	(0.01326)
formal × age=1	-0.00130	-0.00130	-0.00202
	(0.01340)	(0.01340)	(0.01341)
formal × age=2	-0.05867	-0.05865	-0.05980
	(0.01971)	(0.01972)	(0.01973)
formal × schooling=1	0.00757	0.00759	0.00705
	(0.01584)	(0.01584)	(0.01584)
formal × schooling=2	-0.01132	-0.01129	-0.01188
	(0.03014)	(0.03015)	(0.03015)
formal × age=1 × schooling=1	-0.03949	-0.03946	-0.03900
	(0.02165)	(0.02166)	(0.02165)
formal × age=1 × schooling=2	-0.04741	-0.04736	-0.04687
	(0.03541)	(0.03542)	(0.03541)
formal × age=2 × schooling=1	0.02215	0.02214	0.02315
	(0.04106)	(0.04106)	(0.04107)
formal × age=2 × schooling=2	-0.03861	-0.03861	-0.03781
	(0.05091)	(0.05091)	(0.05091)
formal × construction	-0.07797	-0.07799	-0.07838
	(0.02362)	(0.02362)	(0.02362)
formal × commerce	-0.03311	-0.03311	-0.03283
	(0.01744)	(0.01744)	(0.01744)
formal × services	-0.05389	-0.05388	-0.05393
	(0.01065)	(0.01065)	(0.01065)
Intercept	0.30441	0.30442	0.30672
	(0.02615)	(0.02615)	(0.02621)
N	242,501	242,501	242,501

Note: log of real wages is the dependent variable. Regressions include *year*, *sector*, and *individual fixed effects*. Students and individuals with schooling transition were excluded.

Table 7: Transitions between Sectors and Self-Employment

5th int →	**Informal**		**Formal**		**Self-Employed**		**Total**
↓ *1st int*	**Count**	**Frequency**	**Count**	**Frequency**	**Count**	**Frequency**	
Informal	13,892	51.24%	7,813	28.82%	5,406	19.94%	27,111
Formal	6,481	8.13%	69,482	87.11%	3,801	4.77%	79,764
Self-employed	5,493	12.72%	3,564	8.25%	34,124	79.03%	43,181
Total	25,866	17.24%	80,859	53.89%	43,331	28.88%	150,056

Table 8: Segmentation by education and life-cycle

Variable	**Coefficient** (Std. Error)								
	age=0		Δ		age=1		Δ		age=2
schooling=0	—				**-0.00130**	←	-0.05737	→	**-0.05867**
					(0.01340)		(0.01823)		(0.01971)
					↑				↑
Δ					-0.03191				0.02973
					(0.01507)				(0.03793)
					↓				↓
schooling=1	**0.00757**	←	-0.04078	→	**-0.03321**	←	0.00427	→	**-0.02894**
	(0.01584)		(0.01832)		(0.01653)		(0.03651)		(0.03605)
					↑				↑
Δ	-0.01890				-0.02683				-0.07966
	(0.03067)				(0.02148)				(0.05089)
					↓				↓
schooling=2	**-0.01132**	←	-0.04871	→	**-0.06003**	←	-0.04857	→	**-0.10860**
	(0.03014)		(0.03338)		(0.02050)		(0.04148)		(0.03931)

Table 9: Segmentation by position in the household

Variable	**Coefficient** (Std. Error) *female*				*male*
non-head	—				**0.03692** (0.01576)
					↑ -0.03900 (0.01323) ↓
head	**0.01441** (0.01547)	←	-0.01649 (0.01405)	→	**-0.00208** (0.01437)
non-spouse	—				**0.03692** (0.01576)
					↑ -0.05663 (0.04803) ↓
spouse	**-0.03727** (0.01526)	←	0.01756 (0.04834)	→	**-0.01971** (0.04857)

Table 10: Segmentation by wage quartiles

Variable		**Coefficient** (Std. Error)
formal × *1st quartile of wage* ($\hat{\beta}_1$)		0.14811 (0.00893)
formal × *2nd quartile of wage* ($\hat{\beta}_2$)		0.07763 (0.00843)
formal × *3rd quartile of wage* ($\hat{\beta}_3$)		0.01278 (0.00905)
formal × *4th quartile of wage* ($\hat{\beta}_4$)		-0.04385 (0.00982)
F-tests		
Hypothesis	Test statistic	
$H_0: \beta_1 = \beta_2$	$\hat{\beta}_1 - \hat{\beta}_2$	0.07047 (0.01220)
$H_0: \beta_2 = \beta_3$	$\hat{\beta}_2 - \hat{\beta}_3$	0.06485 (0.01232)
$H_0: \beta_3 = \beta_4$	$\hat{\beta}_3 - \hat{\beta}_4$	0.05663 (0.01335)

Note: log of real wages is the dependent variable. Regressions include *year*, *sector*, and *individual fixed effects*. Students and individuals with schooling transition were excluded.

Table 11: Extensions

Variable	**Coefficient** (Std. Error)		
	(1)	(2)	(3)
formal	0.07801 (0.00409)	0.11556 (0.04582)	0.08313 (0.00484)
self-employed	–	–	0.09517 (0.00539)
correction term	–	-0.03274 (0.03931)	–
	1st stage: probit dependent variable: formal		
frac	–	0.21798 (0.04527)	–

Note: log of real wages is the dependent variable. Regressions include *year*, *sector*, and *individual fixed effects*. Students and individuals with schooling transition were excluded.