

Modelling Formal vs. Informal Employment and Earnings: Micro-econometric Evidence for Brazil^{*}

Francisco Galvão Carneiro[†] and Andrew Henley[‡]

Resumo

Em vários países em desenvolvimento, especialmente na América Latina, tem-se observado um crescimento estrutural do setor informal, o que parece ser inconsistente com visões tradicionais de que o setor informal se comporta como um setor residual ou que representa um sintoma de segmentação no mercado de trabalho. Este artigo apresenta evidência microeconômica para o caso do Brasil que desafia essa visão convencional mostrando que a distribuição dos rendimentos no setor informal estende-se até os níveis mais elevados da distribuição total de rendimentos. Estima-se um modelo de escolha de setor usando os rendimentos previstos e corrigidos para o viés de seletividade para cada indivíduo nos setores formal e informal. Os resultados desta estimação mostram que os maiores diferenciais de rendimentos entre os setores informal e formal estão associados com uma maior probabilidade de um indivíduo se encontrar no setor informal. Outras características associadas com a escolha do setor informal são o tamanho do estabelecimento, método de pagamento, sindicalização e posição dentro da família. Concluimos que o emprego informal pode ser uma forma desejada de status no setor informal para muitos trabalhadores na América Latina, ao invés de ser a consequência de uma segmentação estrutural ou deslocamento cíclico.

Palavras Chave: Setor Informal, Rendimentos, Viés de Seletividade

Classificação no JEL: J21, J42

Área 06: Economia do Trabalho

Abstract

Many developing economies, especially in Latin America, appear to be experiencing structural growth in the size of their informal sectors, inconsistent with traditional views that the informal sector acts as a buffer against unemployment, or is symptomatic of segmentation in the labour market. This paper presents micro-econometric evidence for Brazil that challenges this conventional wisdom by showing that the distribution of earnings in the informal sector extends well into the upper range of the full distribution. A model of informal sector choice is estimated using selectivity-corrected predicted earnings for each individual in both informal and formal sectors. The results of this show that a higher predicted earnings differential between the informal and formal sector is associated with a greater probability of a worker being employed in the informal sector. Other characteristics associated with informal sector choice are establishment size, payment method, union status and position within the household. We conclude that informal employment may be a desirable form of labour market status for many in Latin America, rather than a consequence of structural segmentation or cyclical displacement.

Keywords: Informal Employment, Earnings, Selection Bias
JEL Code: J21, J42 - Área 06: Economia do Trabalho

* We should like to thank Reza Arabsheibani for help and advice from an early stage in this research, and Hedley Rees for access to programme code.

† Director, Graduate Programme in Economics, Universidade Catolica de Brasilia, UCB SGAN 916, Modulo B, Asa Norte, Brasilia-DF 70.190-160, Brazil;
email: carneiro@pos.ucb.br

‡ Professor of Economics, School of Management and Business, University of Wales Aberystwyth, Cledwyn Building, Penglais, Aberystwyth, Ceredigion, SY23 3DD, Wales, UK;
email: Andrew.Henley@aber.ac.uk

1. Introduction

Informal employment, which is outside the protection of any state employment and social protection system, is endemic in many developing and newly industrialising economies. An established view in the literature is that the informal sector acts as a buffer between formal employment and open unemployment and, consequently, as formal sector jobs are destroyed during restructuring or recession, informal employment rises (Tokman 1992, Portes and Shaunffler 1993). Another traditional way of looking into the issue of why informal labour markets exist is by considering the hypothesis of labour market segmentation. In the presence of segmentation rewards in the different sectors of the economy differ for workers of equal potential productivity and, in this context, the existence of a dual labour market would be consistent with dualism in earnings as long as workers with similar characteristics are paid different wages according to the sector in which they work (Dickens and Lang 1985). In both cases, the informal sector represents the disadvantaged sector in a dualistic labour market.

However these traditional views fail to offer much insight into the structural growth in the size of the informal sector. Recent work for a number of developing countries suggests that informal employment may in fact be a desirable choice for workers, because formal sector protection is often inefficient and therefore ineffective, and refutes a traditional segmented, non-competing groups understanding of the relationship between the two sectors (Maloney 1999, Funkhouser 1996, Marcoullier et al. 1997, and Saavedra and Chong 1999). Assuming a standard neoclassical labour market, earnings and occupational choice would only reflect different amounts of inherited characteristics and of investment in human capital. Thus, if workers in the informal sector are paid lower wages, this is because they are less educated or less experienced and/or their human capital endowment might have had an influence in their occupational choice, and not because the two sectors present different wage determination processes. In this case, the labour market for relatively unskilled workers may be seen as integrated with the formal sector, offering desirable jobs with distinct characteristics from which workers may choose from with little queuing (Maloney 1999).

This paper uses recent Brazilian micro data to investigate the determinants of the selection of workers into formal or informal employment, and the relative impacts of different labour market factors on earnings in the two states. The aims of the paper are firstly to assess the relative determinants of earnings, especially human capital, in the two sectors; secondly to investigate the scale of sample selection bias on earnings estimates from ignoring the formal/informal distinction; and thirdly to analyse the determinants of informal employment status and in particular the possible influence of relative predicted informal/formal earnings for the individual on this status selection. On this last issue there would appear to be no previous research at the micro-econometric level.

The remainder of the paper is structured as follows. Section 2 sets out the context for the study, reviewing previous micro and macro empirical work on informal employment. Section 3 discusses the data source used in the present study. Section 4 sets out our econometric methodology. Section 5 presents results for both comparative earnings functions for formal and informal employment and for a selection equation. Both reduced and structural form estimates of a model of informal/formal selection are presented. Section 6 draws conclusions.

2. Structural Influences on Informal Employment

Large informal sectors play an increasingly important economic role in many developing economies, albeit one that is difficult to quantify precisely. Estimates for developing countries show that informality can accommodate up to 50% of the economically active population (Funkhouser 1996). Estimates vary because there is no firm agreement on how to measure the size of the informal sector. Some commentators argue that all self-employed workers should be regarded as in the informal sector, others focus on the numbers who are working but not paying social security contributions and therefore fall outside any systems of employment and social protection. In the Brazilian context a useful indicator of informal employment status is the absence of a signed labour card (“carteira de trabalho assinada”), and this is the empirical definition used in the present paper. The growth in employment of this form has been rapid in Brazil, whilst numbers in formal employment have remained fairly static. For an economically active population of 18 million in 2000, 4.5 million were employed in the informal sector. The relative growth in the informal sector has been rapid from 17% of the economically active population in 1990 to 27% in 2000, and this growth appears to be structural rather than cyclical (Carneiro 1997).

There has been little or no formal investigation of the possible explanations for this growth. However the informal sector appears to have a growing depressant influence on formal sector wage determination in Brazil (Carneiro and Henley 1998). Do workers choose informal employment or are they selected into the informal sector because of the relative abundance of such jobs? Recent work for a number of developing countries suggests that informal employment may in fact be a desirable choice (Maloney 1999, Marcoullier et al. 1997). Indeed this is reinforced by Pradhan and van Soest (1997), also researching in a Latin American context, who find evidence of substantial household labour supply substitution between the two sectors. Low formal sector wages do in fact encourage participation in the informal sector.

Fernandes (1996) estimated logit models for Brazil to assess the probability of a worker being employed in the formal sector using household data for 1989 and found that the probability of being formal increases with age, reaches a peak, and then declines. The author also found that the higher the schooling the higher the probability of a worker being in the formal sector. He also showed that the maximum probability of finding a formal job is achieved at the age of 42 and that men are more likely to be in the formal sector. A similar pattern has been found for a group of Central American countries by Funkhouser (1996) and for Peru by Saavedra and Chong (1999), which reported a significant negative relationship between level of education and informal employment. This evidence is suggestive that an increase in education is associated with a decrease in the probability of informal employment.

Sedlacek et al. (1990) investigated the case of Brazil and reported that unpaid workers move disproportionately into the informal sector, which is suggestive that while in school and just after completing school, many individuals help out at the family business and eventually get paid. They spend on average only two years doing this before moving onto other paid work. Even if this pattern of graduation from school to unpaid work to informal salaried work to other models of work represents the queuing that the dualistic literature might predict, the time spent in informal salaried work is not very long, being similar to the pattern observed in industrialised countries. Maloney (1999), for example, reports that median tenure of young workers (16 to 24 years old) in the United States is only 1.4 years and 3.4 years for workers 25 to 34 years old.

As for the earnings differentials, Fernandes (1996) estimated earnings equations for the two sector and then assessed the determinants of the difference in earnings between workers in the formal and the informal sectors. His estimates indicated lower differentials for men in relation to women and that the differentials tend to grow with schooling at a rate of 2.7% per additional year of schooling. Earlier evidence confirms that earnings differentials are significant between the formal and informal sectors, ranging from 30% in the early 1980s to 45% in the early 1990s in favour of formal employment (Pero 1992, and Cacciamali and Fernandes 1993). However these results must be treated with considerable caution because of the absence of any correction for sample selectivity bias.

3. Data Source

The present paper uses data drawn from the 1997 Brazilian household survey (Pesquisa Nacional por Amostra de Domicílios, PNAD). The PNADs are a series of nationally representative household surveys conducted more or less annually since 1976, using a consistent methodology by the Instituto Brasileiro de Geografia e Estatística (IBGE). All members of each household over ten years in age are asked detailed questions concerning their labour market activity during (in the 1997 case) the week 21st to 27th September. The 1997 PNAD surveyed a total of 109,541 households, comprising 121,078 economically active individuals between the ages of 18 and 65. Of the 73,896 who gave information on whether they had a signed labour card or not, 39.9% were working without one. It seems likely that this proportion is an understatement, given that the high level of non-response to this question may indicate that a large number of respondents did not know whether they had a labour card or where reluctant to reveal that they did not have one.

Table 1 reports descriptive statistics on the two groups for the sample used for estimation purposes.¹ Hourly earnings are defined as reported monthly earnings divided by 4.33 times reported weekly hours of work. Average log hourly earnings for formal workers are over twice those for informal workers (formal: R\$2.01 per hour; informal: R\$1.31 per hour). Figure 1 plots the log earnings distribution for both groups – with and without signed labour cards. Although mean log earnings for informal workers are lower dispersion of earnings is only slightly less in the informal sector. Returning to Table 1 we see that the two groups have similar average ages and tenure. However, informal workers are considerably less likely to have achieved secondary or higher level education, and, although the proportions are very small, more likely to be illiterate. Informal workers are somewhat more likely to be male, and substantially more likely to be non-white. One of the most pronounced differences between the two groups is to be found in the proportions employed in establishments with more than 10 employees. Informal workers are less than half as likely to be working in larger establishments. There are also some substantial and important differences between the groups in terms of their means of payment. While salaried (waged) employment is the dominant form of payment for both groups (91.5% of formal workers, 72.1% of informal workers), informal workers are six times more likely to be paid on commission or according to amount produced, and over 20 times more likely to be paid on piece rates. In fact piece rate payment for formal workers is extremely rare. Finally average other household earnings between the two groups are very different, both groups come from households with similar average other income (that is labour income from other household

¹ The sample is further reduced to 41,034 by the absence of data on other covariates, including earnings, used in the empirical analysis.

members, or other non-labour income). Although the proportions of the sample with multiple jobs are low (below 4%), informal workers are somewhat more likely to hold more than one job.²

4. Modelling the Selection of Workers between Formal and Informal Employment

In order to model the selection between the formal and informal sectors, we adopt what is now a conventional approach to the simultaneous modelling of a participation decision and earnings. This entails a three-stage estimation process.³ In stage one we estimate a reduced-form probit model of the formal vs. informal decision. This is used to construct a sample selection correction term that, in stage two, is incorporated into conventional Mincerian semi-log earnings functions for the formally employed and informally employed. This allows us to control for any comparative earnings advantage or disadvantage that the two groups have in their respective sectors, and correct for the bias that this sample selection effect might impart to the determinants of earnings in each state. In stage three the earnings functions results are used to construct a predicted earnings differential between the formal and informal sector for each individual. This is included in a second probit estimation of the structural model of the participation decision.

The model proceeds from the assumption that an individual will rationally be allocated to formal employment if the utility derived from that allocation exceeds the alternative available from informal employment. We assume that the difference in utility from the two sectors is a linear function of the difference in (log) earnings (Y) and a vector of individual characteristics (X). Thus the probability that the difference between utility from self-employment and from employment being greater than zero is:

$$\Pr(U_i^I - U_i^F \geq 0) = \Pr(\alpha + \beta(\log Y_i^I - \log Y_i^F) + \gamma X_i + \varepsilon_i \geq 0) \quad (1)$$

where U_i^I and U_i^F are indices of utility derived by individual i from states of informal employment, I , and formal employment, F . α is an intercept and β and γ are coefficient vectors. Under the assumption that ε_i is an error term distributed $N(0, \sigma^2)$, then (1) can be estimated as a probit model (the “structural” probit). However earnings are only observed in one of the two states, so a Heckman (1979) two-stage procedure must be used to construct predicted earnings for each individual in each state.

Define an indicator variable H as follows: $H_i = 1$ if $U_i^I - U_i^F \geq 0$ and 0 otherwise. Thus $Y_i = Y_i^I$ when $H_i = 1$ and $Y_i = Y_i^F$ when $H_i = 0$. We assume that earnings are determined according to semi-log earnings functions as follows:

$$\log Y_i^I = \delta^I Z_i^I + e_i^I \quad (2)$$

and

$$\log Y_i^F = \delta^F Z_i^F + e_i^F \quad (3)$$

² Where an individual reports multiple jobs, the data in the analysis refer to the main job (as defined by the respondent).

³ This methodology is developed by Lee (1978) in the context of the union participation decision, and has been adopted, for example, for self-employment choice by Rees and Shah (1986)

where Z are covariate vectors, δ are coefficient vectors, and e are random errors distributed $N(0, \sigma^2)$ and $N(0, \sigma^{F2})$ respectively. The model is identified by the exclusion from Z of elements of X . It is well known under the circumstances presented here that estimation of (2) and (3) by OLS will be inconsistent. Consistent estimates can be obtained by estimating, as a probit, a reduced form of equation (1):

$$\Pr(U_i^I - U_i^F \geq 0) = \Pr(\pi^1 X_i + \pi^2 Z_i + \varepsilon_i) = \Pr(\psi_i + \varepsilon_i) \geq 0 \quad (4)$$

These estimates are used to construct a selectivity correction term (inverse Mills ratio) that is incorporated into the earnings functions as follows:

$$\text{and } \log Y_i^I = \delta^I Z_i^I + \lambda^I \frac{\phi(\psi_i)}{\Phi(\psi_i)} + e_i^I \quad (5)$$

$$\log Y_i^F = \delta^F Z_i^F + \lambda^F \frac{\phi(\psi_i)}{1 - \Phi(\psi_i)} + e_i^F \quad (6)$$

where λ are coefficients and ϕ and Φ are the density and cumulative distribution functions of a standard normal variable respectively. (5) and (6) are used to construct predicted earnings in each state for all individuals and which in turn are used to replace Y^I and Y^F in the estimation of equation (1).

5. Results

Table 2 reports estimation results for the reduced form selection equation (equation 4) and the selectivity corrected earnings functions for formal and informal workers (equations 5 and 6). Both selection and earnings equations include 5-segment spline functions in age and 8-segment spline functions tenure with the employer⁴, as well as schooling dummy variables (with elementary schooling being the reference state) and dummy variables for gender, ethnicity, union status and an urban address. Industrial, occupational and regional dummy variables are also included but for reasons of space their coefficients are not reported. The selection equation is additionally identified through the inclusion of variables for establishment size, position in the household, payment method, other household income and multiple job holding. In most cases these variables are strongly statistically significant in the selection equation. We will turn to detailed discussion of the influences on informal/formal selection when we discuss our structural equation estimates shortly.

Turning to the results for the earnings equations, we obtain a better overall equation fit for the formal workers sample than for the informal sample. We find significant linear age effects for both groups of a similar order of magnitude. The absence of significant coefficients for the higher spline segments points to no evidence of non-linearity in the age-earning profile for either group⁵. Earnings for both groups appear to be decreasing in tenure, but in particular for formal sector workers, whose earnings fall at a rate of 1.6 per cent per annum for the first five years with an employer and 3.5 per cent per annum between five and ten years, with a

⁴ Semi-parametric spline functions are used in preference to quadratic polynomials in age and tenure as they are less restrictive on functional form.

⁵ This is consistent with earlier findings by Fernandes (1996) who reports that the coefficients of nonlinearities in the age variable are not statistically significant.

further significant increase in the decline after 20 years. For informal workers, earnings begin to decline significantly after five years at nearly 4 per cent per annum⁶.

Higher levels of schooling at secondary level or above significantly raise earnings relative to the reference category for formal workers, and at college level or higher for informal workers. The relative rates of return for formal workers for secondary and college education are substantially higher than for informal workers. Secondary schooling raises earnings relative to an elementary school graduate by 11 per cent for formal workers, but only by (a statistically insignificant) 4.5 per cent for informal workers. For a college education the relative rates are 62% and 49% per cent respectively, with an even wider gap for those very few workers with post-graduate qualifications.

Being female has a relatively bigger impact on earnings for formal workers (-35%) than for informal workers (-27%). However, the relative premium for being white is the nearly the same in both groups (13-14%). The union differential is significant for both groups but slightly larger for informal workers at just over 16%. Finally urban informal workers do relatively better than their formal counterparts.

The “lambda” coefficient in the earnings equations reports the estimate for selectivity correction term in equations (5) and (6). In both cases it is statistically significant and positively signed, showing that both informal and formal workers enjoy a comparative earnings advantage in the groups into which they have been selected. The “rho” coefficient reports the degree of correlation between the error terms in the selection equation and the relevant earnings equation.

From the estimated earnings equations in Table 2 we are able to generate predicted earnings in each state for each sample individual,⁷ in order to construct a predicted wage differential. This is then included in the structural probit, and the results for this are presented in Table 3. The coefficient on the predicted wage differential attracts a negative and strongly significant coefficient, implying, as would be expected, that the smaller the difference between available earnings in the informal and formal sectors the more likely an individual is to be working in the informal sector. A 1 per cent increase in the ratio of the predicted informal wage to the predicted formal wage raises the probability that an individual is working in the informal sector by 5.1 percentage points.

The results also show that female, white, unionised, urban workers, workers in large establishments and heads of household are significantly less likely to be employed in the informal sector, with the biggest quantitative effect showing for union status. Being female reduces the probability of being in informal employment by 13.4 percentage points, being white by 6.4 percentage points, being unionised by 23 percentage points and being and urban worker by 13 percentage points. A worker in an establishment with more than 10 employees is 33 percentage points less likely to have informal status. A head of household is 3.6

⁶ This result could be reflecting a situation in Brazilian labour markets that is characterised by high labour turnover rates. According to Ramos and Carneiro (1997), Brazilian workers have a very high intertemporal discount rate and prefer to change jobs and receive severance payment rather than remaining employed in a job that will not lead them to promotions and/or improve their human capital endowment.

⁷ A correction factor equal to $0.5\sigma_j^2$ where $j = I, F$ as appropriate is added to predicted earnings in each state to allow for the fact that the earnings function has semi-log form.

percentage points less likely to have informal status, and a spouse is 1.4 percentage points less (significant at 6%).

Payment method also continues to have a statistically significant, quantitatively large impact on the probability of informal status in the structural model. Being paid by commission or output raises the probability of being informal by 52 percentage points, and by piece rate by 59 percentage points. Having more than one job raises the probability of informal status in the main job by a statistically significant 8 percentage points. Finally an extra 1000 Reais per month of other household income raises the probability of having informal status by a statistically significant 1.6 percentage points. This last result is perhaps surprising given that additional income from other sources in the household might reduce the marginal utility of the opportunity cost of paying social insurance contributions. However it may simply indicate that a worker feels better protected against income risk if there is other household income and the need for social and employment protection is attenuated.

The structural probit regression reported in Table 3 also included industry and occupational dummy variables. These are jointly statistically significant and many of the individual coefficients, especially for occupation, are individually statistically significant. The largest positive quantitative impacts on the probability of informal status arise in construction and service industries and in service occupations. A number of occupations are associated with significantly reduced probability of informal status, particularly administrative occupations, but also manufacturing, retailing, and transport and communications occupations.

6. Conclusions and Assessment

This paper has presented micro-econometric estimates of the determinants of earnings and selection between informal and formal status, accounting for the simultaneous determination of the two, using recent Brazilian data. Unsurprisingly we find that returns to schooling are lower in the informal sector. Our results show that formal/informal sample selection bias is quantitatively important in modelling earnings and has a statistically significant impact on earnings for both groups. Both formal and informal workers appear to enjoy comparative earnings advantage in their respective sectors, suggesting that, or at least consistent with, selection between the two states being a rational choice. This conclusion is reinforced in our structural model of the selection decision where the earnings differential between informal and formal workers has a strong, statistically significant effect on the probability of working in the informal sector. To this extent our results support the recent conclusion of Marcoullier et al. (1997), Saavedra and Chong (1999) and Maloney (1999) that informal employment may be a desirable form of labour market status for workers in Latin American economies, rather than informal employment arising from structural segmentation of the labour market into two non-competing groups.

As to the explanations for the trend growth in the size of the informal sector in Brazil, our approach suggests that comparative cross-sectional research, modelling selection at different points in time, particular both before and after the Brazilian trade liberalisation of the early 1990s, will be a fruitful avenue for future work. For the moment our results suggest that some of the strongest influences might arise from industrial restructuring away from manufacturing towards services, from changes in unionisation, and from changes in methods of payment from traditional waged employment towards pay related to worker performance or productivity.

References

- Cacciamali, M.C. and Fernandes, R. (1993), "Distribuição dos trabalhadores e diferenciais de salários entre os mercados de trabalho regulamentado e não-regulamentado", *Pesquisa e Planejamento Econômico*, 23(1):135-56.
- Carneiro, F.G. (1997), "The changing informal labour market in Brazil: cyclicity versus excessive intervention", *Labour*, 11(1): 3-22.
- Carneiro, F.G. and Henley, A. (1998), "Wage determination in Brazil: the growth of union bargaining power and informal employment", *Journal of Development Studies*, 34(4): 117-138.
- Dickens, W.T. and Lang, K. (1985), "A test of dual labour market theory", *American Economic Review*, 75(4):1-22.
- Fernandes, R. (1996), "Mercado de trabalho não-regulamentado: Participação relativa e diferenciais de salários", *Pesquisa e Planejamento Econômico*, 26(3):417-41.
- Funkhouser, E. (1996), "The urban informal sector in Central America: Household survey evidence", *World Development*, 24(11):1737-51.
- Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47:475-92.
- Lee, L-F (1978), "Unionism and wage rates: a simultaneous equation model with qualitative and limited dependent variables", *International Economic Review*, 19: 415-33.
- Maloney, W.F. (1999), "Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico", *World Bank Economic Review*, 13(2): 275-302.
- Marcoullier, D., Ruiz de Casilla, V. and Woodruff, C. (1997), "Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, 45(2): 367-92.
- Pero, V.L. (1992), "A carteira de trabalho no mercado de trabalho metropolitano brasileiro", *Pesquisa e Planejamento Econômico*, 22(2):305-402.
- Portes, A. and Schaunffler, R. (1993), "Competing perspectives on the Latin American informal sector", *Population and Development Review*, 19(1):33-60.
- Pradhan, M. and van Soest, A. (1997), "Household labor supply in urban areas of Bolivia", *Review of Economics and Statistics*, 79(2): 300-310.
- Ramos, C.A. and Carneiro, F.G. (1997), "Rotatividade e instituições: Os benefícios aos trabalhadores desligados incentivam os afastamentos?", Discussion Paper N. 503, IPEA.
- Rees, H. and Shah, A. (1986), "An empirical analysis of self-employment in the UK", *Journal of Applied Econometrics*, 1:95-108.

Saavedra, J. and Chong, A. (1999), "Structural reform, institutions and earnings: Evidence from the formal and informal sectors in urban Peru", *Journal of Development Studies*, 35(4):95-116.

Sedlacek, G.L., Barros, R.P. and Varandas, S. (1990), "Segmentação e mobilidade no mercado de trabalho: A carteira de trabalho em São Paulo", *Pesquisa e Planejamento Econômico*, 20(1):87-103.

Tokman, V. (1992), "The informal sector in Latin America: From underground to legality", in Tokman, V. (ed.), *Beyond regulation: The informal economy in Latin America*, International Labour Office, Lynne Reinner, pp. 3-22.

Table 1: Sample Characteristics

	Formal Workers (with signed labour card)		Informal Workers (without signed labour card)	
	Mean	Std. Dev.	Mean	Std. Dev.
Log (hourly earnings)	0.6991	0.7936	0.2721	0.8353
Age (years)	33.3484	9.9039	31.9967	10.9345
Tenure (years)	18.5861	11.0922	18.6063	11.7860
Highest Education Level:				
Elementary schooling	0.1463		0.1612	
Primary schooling	0.4395		0.5649	
Secondary schooling	0.3161		0.2160	
College degree	0.0907		0.0511	
Masters/Doctorate	0.0038		0.0014	
Illiterate (no schooling)	0.0036		0.0054	
Female	0.3252		0.2685	
White	0.5900		0.4972	
Union member	0.2942		0.0449	
Urban	0.9494		0.9258	
Employer >10 employees	0.7689		0.3270	
Head of household	0.5393		0.4683	
Spouse of head of household	0.1506		0.1289	
Paid by commission	0.0227		0.1856	
Piece rate	0.0023		0.0480	
Daily contract	0.0591		0.0425	
Other contract	0.0005		0.0031	
Other household income, R\$ per month	622.87	1018.33	605.37	1457.11
>1 job held	0.0308		0.0350	
N	29967		11067	

Table 2: Selectivity Corrected Estimates of Earnings Functions for Formal and Informal Employment

	Reduced Form Selection Probit Pr(informal)		Formal Workers Log (Hourly Earnings)		Informal Workers Log (Hourly Earnings)	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Age	-0.1259	0.0080**	0.0692	0.0038**	0.0739	0.0060**
Age >25 (marginal effect)	0.0722	0.0088**	-0.0001	0.0039	0.0075	0.0068
Age >35 (marginal effect)	0.0269	0.0079**	0.0047	0.0031	-0.0096	0.0068
Age >45 (marginal effect)	0.0315	0.0105**	-0.0028	0.0043	0.0134	0.0088
Age >55 (marginal effect)	0.0210	0.0150	-0.0018	0.0068	-0.0117	0.0118
Tenure	0.0795	0.0153**	-0.0160	0.0067**	-0.0011	0.0129
Tenure >5 (marg. effect)	-0.0152	0.0195	-0.0191	0.0083**	-0.0385	0.0166**
Tenure >10 (marg. effect)	-0.0011	0.0143	-0.0085	0.0059	-0.0083	0.0116
Tenure >15 (marg. effect)	-0.0156	0.0143	-0.0049	0.0057	-0.0154	0.0119
Tenure >20 (marg. effect)	-0.0130	0.0159	-0.0138	0.0063**	-0.0123	0.0133
Tenure >25 (marg. effect)	-0.0197	0.0176	0.0097	0.0070	0.0133	0.0147
Tenure >30 (marg. effect)	0.0197	0.0194	-0.0140	0.0078	-0.0043	0.0162
Tenure >35 (marg. effect)	-0.0253	0.0152*	-0.0036	0.0064	-0.0090	0.0124
Primary schooling	-0.0086	0.0382	0.0192	0.0159	0.0015	0.0314
Secondary schooling	0.0304	0.0522	0.1097	0.0219**	0.0445	0.0425
College degree	0.3443	0.0718**	0.6241	0.0298**	0.4860	0.0594**
Masters/Doctorate	0.0373	0.1869	1.1388	0.0597**	0.7835	0.1873**
Illiterate	-0.1645	0.1165	0.1960	0.0544**	0.0910	0.0889
Female	-0.1020	0.0225**	-0.3450	0.0077**	-0.2710	0.0166**
White	-0.0133	0.0174	0.1324	0.0072**	0.1370	0.0141**
Union member	-0.9384	0.0266**	0.1487	0.0082**	0.1620	0.0331**
Urban	-0.1450	0.0327**	0.0891	0.0148**	0.1281	0.0249**
Employer >10 employees	-1.0019	0.0164**				
Head of household	-0.1096	0.0219**				
Spouse of head of hh	0.0119	0.0288				
Paid by commission	1.3984	0.0306**				
Piece rate	1.5883	0.0782**				
Daily contract	0.0256	0.0345				
Other contract	1.2918	0.2062**				
Other hh income R\$'000	0.0390	0.0085**				
>1 job held	0.2244	0.0448**				
Intercept	2.4242	0.1610**	-0.9172	0.0756**	-1.4033	0.1231**
Lambda			0.0611	0.0141**	0.0746	0.0152**
Rho			0.1106		0.1113	
N	41034		29967		11067	
LogL	-16795.8					
LR Chi-sq (51)	14251.1**					
R-squared			0.5034		0.3605	

Notes: All equations also include 2-digit industry dummies (9), 2-digit occupation dummies (7) and regional dummies (4).

* denotes coefficient significant at 10%, ** at 5%.

Table 3: Structural Probit Estimates for Informal vs Formal Status

	Coefficient	Standard Error	Marginal Effect
Predicted $\log(W^I) - \log(W^F)$	5.1303	0.1772**	1.4476
Female	-0.5207	0.0254**	-0.1344
White	-0.2231	0.0163**	-0.0637
Union member	-1.0486	0.0267**	-0.2260
Urban	-0.3976	0.0332**	-0.1274
Employer >10 employees	-1.0545	0.0165**	-0.3267
Head of household	-0.1276	0.0186**	-0.0361
Spouse of head of hh	-0.0512	0.0272*	-0.0142
Paid by commission	1.4515	0.0307**	0.5229
Piece rate	1.6539	0.0790**	0.5915
Daily contract	0.0093	0.0346	0.0026
Other contract	1.3310	0.2102**	0.4883
Other hh income R\$'000	0.0559	0.0083**	0.0158
>1 job held	0.2590	0.0447**	0.0801
Intercept	2.4328	0.0952	
N	41034		
Log L	-16810.2		
LR Chi-sq (30)	14222.2**		

Notes: Equation also includes 2-digit industry dummies (9) and 2-digit occupation dummies (7). Marginal effects show the effect of a discrete change in the case of dummy variables.

* denotes coefficient significant at 10%, ** at 5%.

Figure 1: Earnings Distributions in the Formal and Informal Sectors

