Regional Variation in Informal Employment:
Skills, Norms, and Governance[*]

November 11, 2009

ERIK JONASSON

OECD, Trade & Agriculture Directorate, Paris, France
and
Lund University, Lund, Sweden

erik.jonasson@nek.lu.se
Telephone: +33 637 832 388 or +46 708 866 300

ABSTRACT

Informal economic activity across countries has been studied thoroughly in the empirical literature, but little research addresses the sources of variation in informality on the sub-national level. This paper analyzes reasons for regional variation in informal employment. It develops a theoretical model, which predicts that worker skill endowment, quality of local governance, and social norms are important determinants of the degree of informal employment in a city. The empirical application draws on data from Brazil, where 45 percent of the urban labor force are employed informally. The degree of such employment, however, varies substantially across regions, with some cities having 20 percent or less informal employment and others having 80 percent or more. The empirical evaluation supports the predictions of the model and shows that informal employment is lower in regions with better governance and stronger social norms for compliance with tax and labor regulation. The analysis also supports the notion of a "skill threshold" for successful entry into the formal sector. Endogeneity concerns are raised and assessed along with other robustness checks of the empirical results.

Key words: informal employment, social norms, government effectiveness, Latin America, Brazil

JEL classification: J21, J24, O17, R23

---

[*] Revision of Chapter 4 of *Labor Markets in Transformation: Case Studies of Latin America,* Lund Economic Studies, No 155.

1. INTRODUCTION

All economies, from the least developed to the most modern and wealthy, contain elements of informal, or unofficial, economic activity. It is estimated that as much as 60 percent of economic activity is unaccounted for in the official records of some countries in Africa. The size of the unofficial economy in Western Europe is believed to range from 10 percent of GDP in Switzerland to almost 30 percent in Greece. About nine percent of the economic activity in the United States is estimated to be unofficial (Schneider, 2005). These numbers are not vastly altered when defining informality as the share of the labor force that does not participate in the formal labor market. With the labor force definition, informality in Latin America ranges from 25 percent in Chile to 75 percent in Paraguay and Bolivia (Perry et al., 2007).

If a large number of market transactions in an economy take place informally, or underground, the tax base will be hollowed out, ultimately jeopardizing the standard of public goods and services (Johnson et al., 1998). For the individual firm, informality tends to imply more costly contract enforcement and limited access to financial capital, constraining its ability to grow (Loyaza, 1996). For the worker, informality to a large extent means being outside social security arrangements and being unprotected by labor regulation (Jütting et al., 2008).

Studies that seek to explain cross-country differences in the relative size of the informal economy contain mixed evidence on what causes one country to have a larger informal sector than the other. Excessive burden of taxes and social security contributions, strict regulations in the official economy, declining loyalty towards public institutions, and declining tax morale are some aspects that are frequently pointed at as driving economic activity underground (Schneider and Enste, 2000). Besides these regulatory and institutional aspects, informality is often found to decrease with the aggregate level of productivity in a country (Loayza and Rigolini, 2006).

The extent to which economic activity takes place informally varies not only across countries, but also across regions within countries. In Brazil, about 45 percent of the urban labor force works in the informal sector. Among its 5 500 municipalities, however, informality ranges between 20 and over 80 percent of the urban labor force. Clearly, tax burden, labor regulation, and other formal institutions that are common nationwide, are not the only factors that determine the extent to which economic activities take place informally.

This study adds to the theoretical and empirical literature on informal employment by proposing, and empirically evaluating, a model that explains regional variation in informal employment, while accommodating several micro-empirical regularities in the informal sector. The model hypothesizes that regional differences in informality are due to differences across regions in worker skill endowment and quality of local governance, including support and implementation of the formal institutional framework that governs market transactions in the formal sector. The model also accounts for non-pecuniary costs and benefits in the informal sector that affect the incentive structure faced by workers. One such hypothesized cost is the moral cost of evading taxes and not complying with labor market regulations.

The theoretical framework is evaluated empirically using data from the Brazilian Demographic Census for the year 2000, combined with municipal information on local governance and public sector service provision. Regional variation in informality is observed in several countries, but Brazil makes a particularly interesting study object. It has one of the strictest labor market regulations in the world (Almeida and Carneiro, 2009). At the same time, the country tends to score low in international

comparisons of the perceived quality of certain institutions and in the level of public trust in politicians. A recent comparison ranks Brazil 122nd and 129th among 134 countries in terms of "public trust in politicians" and "wastefulness of government spending", respectively (World Economic Forum, 2008). This suggests that, while the formal political ambitions are high in terms of willingness to provide its citizens *Ordem e Progresso*, implementation and enforcement by local governments vary substantively across different regions. The size of Brazil in terms of area and population gives room for large differences in socioeconomic and cultural characteristics across regions, which is likely to reinforce regional differences in policy implementation.

The empirical results support the predictions of the theoretical model. In particular, human capital level, quality of local governance, and strength of social norms are all related negatively to the size of the informal sector. These results stand up to a series of robustness checks, but given the cross-sectional nature of the data, limitations of the empirical assessment still need to be acknowledged.

2. PREVIOUS EMPIRICAL FINDINGS

There is a fairly large empirical literature that evaluates the effects of worker and employer characteristics on the probability of participating in the informal sector. This literature generally does not test for regional or institutional effects, or for the potential role of local governance or public goods provision. Data limitations tend to prevent such analysis on a disaggregated level beyond the inclusion of regional dummy variables (El Badoui et al., 2008; Gong et al., 2004). Since personal and firm characteristics clearly matter for sector participation, existent micro-empirical evidence constitutes important guidance in theoretical as well as empirical analysis of non-individual-specific determinants of informal employment.

Micro-level studies generally conclude that the probability of working in the informal sector decreases with human capital endowment measured by years of education, job tenure, and experience (Funkhouser, 1996; Marcouiller et al., 1997; Gong et al., 2004). In terms of age, it tends to be the youngest and the oldest in the labor force who participate in the informal sector. The young mainly work as informal labor, whereas the older become self-employed (Perry et al., 2007). These findings are consistent with the "threshold view" that the worker needs a certain level of skill to benefit from participation in (or to get access to) the formal sector, as opposed to working in the informal sector. There is some evidence that position in the household, household size, and the presence of children in the household affect sector choice. This probably captures a gender effect. Women with children and married women tend to be more likely to participate in the informal sector (Funkhouser, 1996; Amuedo-Dorantes, 2004). The reason could be that the informal sector offers a higher degree of flexibility in terms of work hours and work location (Maloney, 2004).

On the regional, within-country, level, Chaudhuri et al. (2006) analyze socioeconomic, political, and institutional factors to explain differences in the size and growth of the shadow economy across Indian states. They find some evidence that the expansion of the shadow economy is less pronounced in states where the press is free and independent (used as a proxy for well-functioning democracy) and where the economy is more liberalized. Torgler and Schneider (2007) find that the shadow economy in Swiss cantons is smaller where tax morale is higher and where direct democratic participation is higher. Empirical evidence for Brazil points at the importance of labor regulation and its enforcement. Almeida and Carneiro (2009) analyze how differences in enforcement of labor regulation across regions in Brazil affect informal employment and unemployment. They find evidence that stricter law enforcement in a

region, measured as the aggregate amount of registration fines issued, leads to less informal employment and to higher unemployment.

Further empirical evidence of the causes of differences in informality across space is, as far as the author is aware, exclusively on a cross-country level. Several studies find that the burden of business regulation is correlated positively with various measures of the informal sector (share of GDP – Johnson et al., 1998, and Friedman et al., 2000; share self-employed in the labor force – Loayza and Rigolini, 2006). Torgler and Schneider (2007), however, find no significant correlation between informality and the burden of labor regulation. The evidence regarding the role of tax rates is mixed. Friedman et al. (2000) find a negative correlation between taxes and informality, even when controlling for GDP per capita. They conclude that high taxes, per se, do not drive businesses underground as long as tax revenues are spent on supplying high-quality public goods and services. Instead, they argue, it is the discretion in the application of rules, and the corresponding corruption that it creates, that cause businesses to exit or avoid the formal economy. Studies that include measures for bureaucratic quality, rule of law, governance and institutional quality all find a negative relationship between these measures and the size of the informal sector (Schneider, 2005).

3. THEORETICAL FRAMEWORK

A static model is developed in this section to explain regional variation in informal employment. The proposed determinants of this variation are regional differences in worker skill distribution, tax rates, and government effectiveness. The latter is understood as the quality of publicly provided goods and services, including the extent to which the local government supports and implements the formal institutional framework of market transactions in the formal sector. These three factors are assumed to affect informality by determining the expected income of the worker. Non-pecuniary effects of informal employment enters the model as a fourth factor by affecting the perceived attractiveness of informal employment, beyond expected income. The model is inspired by Loayza and Rigolini (2006), Galiani and Weinschelbaum (2007), and to a lesser extent by Rauch (1991). Sector of employment is modeled as a choice by the individual, rather than exogenously determined by segmented labor market mechanisms. There is an ongoing debate in the literature about which characterization of the informal sector is more realistic (Perry et al., 2007; Maloney, 2004). The implicit assumption made here is that the employment outcome is generally a choice, but within a limited opportunity set. In some cases the worker's opportunity set might be reduced to what practically becomes only one option; employment in the informal sector.

The model assumes that workers differ by skill endowment s which is distributed among workers according to some density function $h(s)$ ($0 \leq s \leq 1$). Skill endowment determines the productivity of the worker, measured in number of homogenous labor units, $l(s)$. A worker can devote labor units to work in the formal sector (*FS*) or the informal sector (*IS*), but not to both. Two cases are characterized below. In Case I, the worker maximizes utility by choosing sector with highest expected income. In Case II, the worker chooses sector with highest expected benefit, by taking into account non-pecuniary costs and benefits of informal employment.

Case I: Comparing expected income

Consider first the case in which the worker only considers expected income when choosing sector of employment. Utility is assumed to increase monotonically in income, so that the worker maximizes utility by choosing to work in the sector that gives the highest expected income. The decision to work is pre-determined outside the model, hence the consumption/leisure decision is abstracted from. The utility maximization problem reduces to:

$$\max_j \ U(y^j), \quad j = \{FS, IS\} \tag{1}$$

where y^j is expected labor income in sector j.

The formal and informal sectors are assumed to differ from each other in several respects, which affect expected income and hence the incentives faced by the worker in the choice of sector. First, the institutional arrangements in the formal economy – whose quality and potential role depend on local government effectiveness – allow for higher labor productivity than in the informal sector. Second, the expected income in the formal sector is affected by the tax rate as well as the probability of finding a job. Third, in the informal sector there is a risk of income loss due to the risk of being charged with tax evasion. These factors are introduced below.

Labor productivity is characterized by positive but diminishing returns to skill in both sectors. For the informal sector, labor productivity, measured in units of homogenous labor, is given by:

$$l^{IS}(s) = s^\alpha \qquad (0 < \alpha < 1) \tag{2}$$

In the formal sector, labor productivity for any level of skill is affected by the quality of the institutional framework that governs market transactions and business practices in the formal sector. The model allows for different interpretations of the manner in which local governance affects labor productivity. A broad interpretation is that, with formal institutions that are well implemented and enforced by the government, firms face better prospects for benefiting from economies of scale through cheaper access to capital and better ability to enforce complex contracts (de Paula and Scheinkman, 2007; Straub, 2005). As a consequence, production in the formal sector will be more capital intensive on average, hence increasing the productivity of labor. A narrower interpretation is that local governments can supply public services and support institutions that directly affect labor productivity, independently of production technology, in an efficiency-wage manner. It could be that the worker is more productive in the formal sector due to an increased sense of transparency, security, and stability in the work environment of the formal sector, compared to the informal sector. If g denotes the effectiveness by which the government is able to maintain this institutional framework, then labor productivity in the formal sector is given by:

$$l^{FS}(s, g) = (1 + g)s^\alpha \quad (-1 < g < 1) \tag{3}$$

The formal and the informal sectors are assumed to be integrated in the sense that one unit of homogenous labor, l, is paid the same in both sectors. For simplicity, assume that the worker faces a perfectly elastic labor demand at a price equal to one. (Self-employed workers do not receive labor income from an employer. These workers get paid for their production of goods and services. For simplicity, assume that production among self-employed is given by $q = l^{IS}(s)$ and that the price of output equals one.)

For workers in the informal sector there is a risk of being charged by the authorities for working informally (or for one's employer acting informally). In case of detection, labor income will be zero. The

risk of being charged, c, is determined by the strength of the enforcement of tax and labor regulation e ($0 < c < 1$; $e > 0$; $c'(e) > 0$). Expected income in the informal sector is given by:

$$y^{IS} = [1 - c(e)]s^{\alpha} \qquad (4)$$

Expected income in the formal sector is affected by the worker's probability of finding a job in that sector and thereby gaining the corresponding productivity enhancement. The worker's perceived probability of obtaining a job is an increasing function of skills. As a simplifying assumption, let the probability be given by:

$$\pi(s) = s^{1-\alpha} \qquad (5)$$

Thus, expected gross income in the formal sector is given by the product of (3) and (5). Imposing labor income tax rate t ($0 < t < 1$), the expected net income in the formal sector is:

$$y^{FS} = (1 - t)(1 + g)s \qquad (6)$$

The threshold level of skill that equates the expected incomes y^{IS} and y^{FS} in the two sectors is:

$$\underline{s} = \left[\frac{(1-t)(1+g)}{(1-c)}\right]^{1/(\alpha-1)} \qquad (7)$$

Thus, whenever the skill level falls short of $\underline{s}$ the individual does not expect to benefit from participating in the formal sector. Given that workers choose the sector rationally (and do not attach utility or disutility to any of the sectors beyond labor income) $y^{IS}(\underline{s}) = y^{FS}(\underline{s})$ denote the highest expected earnings in the informal sector and the lowest in the formal sector, respectively, for given levels of t, g and c. The existence of a formal sector requires that $\underline{s}$ is less than 1, or:

$$\gamma \equiv \frac{(1-t)(1+g)}{(1-c)} > 1 \qquad (8)$$

The γ denotes the ratio of formal to informal income for the highest educated (for whom $s = 1$). If this condition is not satisfied, then the tax rate is too high, governance and institutional quality too low, or the risk of income loss in the informal sector too low for anyone to benefit from working in the formal sector.

The higher the skill threshold, the higher the probability that any given worker will prefer to work in the informal sector. Differentiating (7) with respect to t, g, and c shows that the propensity to work informally increases with t and decreases with g and c. Figure 1 provides an illustration of these effects. For every solution to (7), there is a skill threshold $\underline{s}_0$ below which workers expect to earn more by being in the informal sector than in the formal sector. This is depicted by the segment of the y_0^{IS} curve that lies above the y_0^{FS} line. The effect of an increase in law enforcement, which increases c, is shown by the lowering of the informal-sector income curve from y_0^{IS} to y_1^{IS}, which decreases the skill threshold from $\underline{s}_0$ to $\underline{s}_1$. On the other hand, an increase in the tax rate or a decline in government effectiveness decreases the slope of the formal-sector income curve from y_0^{FS} to y_1^{FS}, which increases the skill threshold from $\underline{s}_0$ to $\underline{s}_2$.

The share of the labor force that works in the informal sector in region m is given by:

$$IS_m = H_m\left[\underline{s}_m(g_m, t_m, e_m)\right] = \int_0^{\underline{s}_m} h_m(s)ds \qquad (9)$$

where $H_m(s)$ is the cumulative density function of skill endowments s in regional m. Hence, differences across regions in the relative size of the informal sector are modeled as regional differences in quality

institutions g, tax rate t, risk of income loss in the informal sector c, but also as differences in the skill distribution in the local labor force, $h_m(s)$. For any given levels of t, g, and c, informality will be higher (lower) in regions with lower (higher) average level of skill.

Case II: Comparing expected benefits

The model outlined above is able to explain why workers in the informal sector tend to have lower skills than those in the formal sector. The higher earnings in the formal sector stem from this skill difference in combination with a productivity-enhancing institutional framework in the formal sector. The model does not predict that there could be an overlap between the sectors in either skills, earnings, or both, i.e. that some workers in the informal sector are more skilled and earn more than some workers in the formal sector. Such an overlap has been found to be extensive in several empirical studies (Pradhan and van Soest, 1995; Funkhouser, 1996; Marcouiller et al., 1997; Bosch et al., 2007) and is evident also in the case of Brazil. One explanation for such a sectoral overlap is the existence of worker-specific non-pecuniary benefits and costs of participation in the informal sector. Non-pecuniary benefits of working in the informal sector could be a higher degree of flexibility in working hours or a greater sense of freedom (Maloney, 2004). A cost could be the moral cost experienced by violating tax or labor legislation (Torgler and Schneider, 2007), as well as a sense of insecurity about one's livelihood, in terms of future earnings, employment contract renewal, or enterprise survival (Jütting et al., 2008).

To take non-pecuniary costs and benefits into account, the utility maximization problem in (1) is adjusted to:

$$\max_j \ U(\hat{y}^j), \quad j = \{FS, IS\} \tag{1'}$$

where $\hat{y}$ denotes total expected benefit of employment. Let $b(x, n)$ denote the net value of the non-pecuniary benefits and costs for the worker of being in the informal sector ($-1 < b < 1$). Assume that b is distributed among workers independently of the skill distribution $h(s)$. The vector x consists of individual-specific characteristics and n represents the strength of the local social norm of obeying formal rules and regulations ($b'(n) < 0$). Assume that b is proportional to income such that the sum of benefits in the informal sector is:

$$\hat{y}^{IS} = [1 + b(x, n)]y^{IS} = [1 + b(x, n)] \times [1 - c(e)]s^\alpha \tag{10}$$

The skill threshold for the worker, which equates total benefits in the informal sector with income in the formal sector, $\hat{s}$, is given by:

$$\hat{s} = \left[\frac{(1-t)(1+g)}{(1-c)(1+b)} \right]^{1/(\alpha-1)} \tag{11}$$

The qualitative effects of changes in t, c, and g are the same as in the case without non-pecuniary benefits. The additional effect to note is the positive effect on $\hat{s}$ of an increase in b. Since $\partial \hat{s}/\partial b$ is positive and $\partial b/\partial n$ is negative, then $\partial \hat{s}/\partial n$ must be negative. Figure 2 illustrates situations in which the worker experiences zero, positive, and negative net benefits in the informal sector. For given levels of t, c, and g, let $\hat{s}_0$ be the skill threshold in the case when the net benefit b is zero. If a worker values the net benefit by $b_1 > 0$, then the total benefit in the informal sector is given by the $\hat{y}^{IS}_1$ curve, which lies above the y^{IS} curve. The $\hat{y}^{IS}_1$ curve intersects the formal sector income curve, y^{FS}, at $\hat{s}_1 > \hat{s}_0$. The worker's monetary income in the informal sector at $\hat{s}_1$, $y^{IS}(\hat{s}_1)$, is lower than the income she expects to earn in the formal sector with the same skill level, $y^{FS}(\hat{s}_1)$. Due to the non-pecuniary benefit, however,

she is indifferent between the two sectors at this skill threshold. If, instead, a worker experiences a negative net benefit $b_2 < 0$ of participating in the informal sector, then the total benefit curve falls below y^{IS} and the skill threshold decreases from $\hat{s}_0$ to $\hat{s}_2$. Despite the fact that $y^{IS}(\hat{s}_2)$ is higher than $y^{FS}(\hat{s}_2)$, the worker is indifferent between the two sectors at this point, due to the disutility attached to work in the informal sector.

Given that b is individual specific, there will be an entire distribution of thresholds $\hat{s}$ within each region. Let $s^* = s - \hat{s}$, so that whenever s^* is below 0, the individual works in the informal sector. If the distribution of s^* is given by $k(s^*)$, the size of the informal sector in region m is given by:

$$IS_m[k_m(s^*)] = \int_{-1}^{0} k_m(s^*)ds^* \qquad (12)$$

4. EMPIRICAL APPROACH

The empirical evaluation of the model is carried out at two levels of analysis. A binomial probability model is first estimated on worker level to assess the extent to which the individual and local factors considered in the model correlate with the probability of having informal employment. A fractional logit model is then estimated on municipal level to assess the extent to which the exogenous variables under consideration also correlate with the municipal size of the informal sector. The methods of the two approaches are discussed below.

The binomial probability model is estimated with a probit model. The difference s^* between the worker's skill level (s) and her participation threshold ($\hat{s}$) enters the probability model as an unobserved latent variable. By assumption, worker i participates in the informal sector if and only if s_i^* is below zero. Thus, if s_i^* is determined by the set of exogenous variables under consideration, then the probability that the individual participates in the informal sector is given by:

$$p_i \equiv prob(is_i = 1) = prob(s_i^* < 0 | s_i,\ x_i,\ r_m,\ z_m) \qquad (13)$$

The binary variable is_i takes the value 1 if the individual works in the informal sector and 0 if in the formal sector. s_i is a set of proxies for worker skills; x_i is a vector of other individual characteristics that are assumed to affect productivity and the individual-specific valuation of the non-pecuniary net benefit in the informal sector; r_m is a set of proxies for government effectiveness and the strength of the social norm in the municipality; and z_m consists of local productivity shifters that might affect the relative productivity in the formal and informal sectors.

The probability model is estimated under the assumption that s_i^* is a linear function of the exogenous variables:

$$s_i^* = s_i\beta_s + x_i\beta_x + r_m\beta_r + z_m\beta_z + u_i \qquad (14)$$

where the βs denote vectors of parameters to be estimated and u_i denotes a residual that is assumed to be normally distributed with zero mean and variance σ^2. Let F_u denote the cumulative distribution function of u. The probit model then gives:

$$\hat{p}_i = 1 - F_u(s_i\hat{\beta}_s + x_i\hat{\beta}_x + r_m\hat{\beta}_r + z_m\hat{\beta}_z) \qquad (15)$$

where $\hat{p}_i$ denotes the estimation of p_i and the $\hat{\beta}$s are the estimated coefficients ($0 \le \hat{p} \le 1$).

On the municipal-level, the relationship between the size of the informal sector and the exogenous variables is estimated using a fractional logit model proposed by Papke and Wooldridge (1996). The share of the labor force in the informal sector in municipality m, IS_m (a fractional variable),

is modeled as a function of aggregates of worker characteristics s_m and x_m, and the same local characteristics, r_m and z_m, as those in the probability model above. The fractional logit approach has some potential advantages over other estimation methods involving a fractional dependent variable, such as the equivalent of the linear probability model or the log-odds transformation. The model has similarities with the regular logit model, with the difference that the dependent binomial variable is replaced with a continuous variable that lies between 0 and 1, and that the estimation is done using a quasi-maximum likelihood procedure. Define X_m as the set of explanatory variables (s_m, x_m, r_m, and z_m). The expected size of the informal sector is assumed to be related with X through a logistic function as:

$$E(IS_m | X_m) = \frac{\exp(X_m \beta)}{1 + \exp(X_m \beta)} \tag{16}$$

Model (17) is then estimated with the Bernoulli log likelihood function:

$$L_m(\beta) = IS_m \times \ln[\,E(IS_m | X_m)\,] + (1 - IS_m) \times \ln(1 - E(IS_m | X_m)) \tag{17}$$

The various robustness checks that have been performed for both models are described separately in the results section.

5. DATA AND DESCRIPTIVE STATISTICS

The main data source used for the empirical analysis is the Brazilian Demographic Census for the year 2000. All worker-level information is derived from these data, and several of the municipal-level variables are aggregated from them as well. The publicly available sample of the Census data includes almost 20 million individual observations, which makes it representative at the level of the municipality. In total, there were 5 507 municipalities in the year 2000, with an average population of about 30 000 people. The Census provides detailed information on employment status, earned income, and a range of socioeconomic variables. To test and control for municipal effects, two data sources are used in addition to aggregates derived from the Census: *Base de Informações Municipais* between 2000 and 2006 and *Perfil dos Municípios Brasileiros - Gestão Pública,* 2005–2006. These databases, provided by the Brazilian Institute of Geography and Statistics (IBGE), contain detailed information on the structure of the local economy, various demographic characteristics and key public sector indicators on the municipal level.

The definition of the informal sector used in this paper is based on the employment categories defined in the Census. Informal employment is defined here as being an unregistered employee, a self-employed person, an unpaid worker, or an employer who has fewer than five employees and does not contribute to any social security institution.[1] Unpaid workers (who constitute about 5 percent of the informal sector) are excluded from the empirical analysis and the descriptive tables. Only the urban labor force is analyzed in this study, mainly because the majority of rural residents are engaged in agriculture and do not have access to either a formal or an informal labor market to the same extent as in urban areas.[2] Using this definition, 45 percent of the urban labor force in Brazil is informal. Approximately 60 percent are employees and about 40 percent are self-employed in the informal sector.

---

[1] Henley et al. (2009) provide an analysis of three alternative measures of informal employment in Brazil, using the PNAD survey. The definition used here corresponds to a large extent to their measure of informality defined as "no signed labor card".

[2] The urban/rural dichotomy in the Demographic Census is determined on an administrative basis and not on a certain population size or density.

The informal sector in Brazil has some of the characteristics that are commonly observed in studies concerned with informal employment in Latin America and elsewhere. Table 1 provides some key indicators. First, labor incomes are on average considerably lower in the informal than in the formal sector. Average earnings per month in the informal sector are just above 400 Reais per month (about 200 US$ in year 2000), compared to 769 Reais in the formal sector. At the 20th percentile of the earnings distribution in the informal sector, earnings are 120 R$ – well below the minimum wage (151 R$) – compared to 221 R$ at the same percentile in the formal sector. The fact that the earnings at the 20th percentile in the formal sector are reached in the informal sector at approximately the 50th percentile shows that there is a considerable overlap in the earnings distributions in the two sectors.

Second, education is lower in the informal sector. On average, a worker in the informal sector has 2.5 years less education than a worker in the formal sector. Similar to the distributions of income, there is an overlap in terms of education between the two sectors. For example, 27 percent of the workers in the informal sector have nine or more years of education, while almost 50 percent in the formal sector have less than nine years of education.

Third, the share of workers who have just entered the labor market or are close to retirement is higher in the informal sector, even though this is not as pronounced as in many other countries (Perry, et al., 2007). One aspect in which Brazil deviates from many other examples is the gender distribution across the sectors. While the informal sector in many countries tends to be over-represented by women (Kucera and Xenogiani, 2009), there is no such gender bias in Brazil. In terms of industrial composition, there is a slight bias towards manufacturing and domestic services in the informal sector compared with the composition of the formal sector.

There is a vast variation of informality in local labor markets. The average of informality on municipal level is 55 percent, but around this mean, informality varies between 20 and 80 percent. There are more than 300 municipalities with shares below 30 percent and over 1 100 with shares above 70 percent. This degree of variation in informality is not notably altered when small municipalities of 10 000 or fewer inhabitants are excluded. Even among the 200 most populated municipalities – each with 400 000 or more inhabitants – informality varies between 25 and 70 percent.

Table 2 gives an overview of the explanatory variables used in the empirical analysis. Human capital is represented by the individual's age and years of education. Squared age and education are included to test for non-linearity. Two additional variables are added to control for labor productivity; an indicator for physical or mental disability, which is likely to affect work ability, and an indicator showing if the worker has recently migrated from a rural area. It could be that people who have recently entered the urban labor force from rural areas face a disadvantage in terms of knowledge of the local economy and have less access to social networks and informal institutions, which affect labor productivity (Fields, 1975; Mazumdar, 1976).

Variables included under the assumption that they affect the utility of non-pecuniary benefits of informal-sector employment are: gender, position in the household, marital status, and indicators for the presence of formal-sector workers and young children in the household. Gender bias in household-related work and responsibilities may imply that women benefit more from the work flexibility in the informal sector than men. This effect may be strengthened if there are children in the household. The employment status by one household member may affect the work incentives by another household member. For example, Maloney (2004) notes that, in the case of Mexico, if one household member has a job in the formal sector, the entire family tends to be covered by several of the fringe benefits of the job. Thus, once one household member has a formal job, there might be little incentive for other household members to take a formal job. Instead they can work in the informal sector for cash income only. This

effect is tested for, in this paper, by variables showing the employment status and income of other household members.

Government effectiveness is tested for by using an index-variable approach. The construction of the index was inspired by the Brazilian *IQIM* index of local institutional quality used by the Brazilian Ministry of Planning as well as the *International Country Risk Guide* produced by the PRS Group. A common feature of these measures is that they are constructed as weighted sums of a range of indicators in order to obtain an index that can rank regions or countries according to quality of governance and institutions.

The composite variable *government effectiveness* (g) is the weighted sum of three indices that are constructed to capture different aspects of local governance and bureaucracy. The first component is *policy formulation* (g_1), which is a proxy for the capacity of the local administration to formulate and implement policy. It is based on 20 indicators, showing to what extent the municipality has councils, development plans, ordinances, and regulations in various domains such as education, urban development, employment, and property ownership. The second component represents *bureaucratic resources* (g_2) and is based on information about: i) employment form of the staff of direct administration (which share has a statutory employment contract and hence benefits from the relatively generous public-sector employment benefits), ii) competence of bureaucrats, measured as the share of the administrative staff who have secondary or higher education, iii) degree of co-ordination between units with different areas of responsibility, and iv) information technology resources. The reason for including employment form is that a bureaucrat with a generous employment contract, which offers a rich set of fringe benefits and employment security, might be more loyal and motivated to do a good job as a civil servant than if he or she has a loose and non-transparent employment contract (Evans and Rauch, 1999). The third component represents the quality of, and access to, *public goods and services* (g_3). This index is based on the teacher/pupil ratio in public primary schools, the number of health centers per municipality inhabitant, the degree of internet services offered to the public, the existence of public libraries, and the degree of support for helping people with housing. All index variables are between zero and one. The correlation coefficients of g_1, g_2, and g_3, range between 0.24 and 0.42. The effect of enforcement of tax and labor regulation (e) is not tested for explicitly since it is difficult to separate this from other qualities of the local bureaucracy (g_2).

To empirically assess the possible role of a social norm influencing the moral cost of acting informally, a variable is needed that is not merely an outcome of informal employment (observed behavior) but is able to capture an attitude affecting the choice of employment. Only a survey that specifically asks for attitudes (such as the international *Latinobarómtro*) would provide a fully satisfactory indicator for this. This information is not available at such disaggregated levels as municipalities in Brazil. Instead, voting participation in municipal and presidential elections (2000 and 2002, respectively) is used as a proxy for such an attitude. The notion of rational ignorance implies that a person will only vote if the expected benefit exceeds the transaction cost of voting. But if there is a strong social norm to vote, and thereby signaling a concern for the municipality or the society as a whole, the decision not to vote could imply social sanctions and a high moral cost (Harbaugh, 1996). It is assumed here that there is some positive correlation between such a voting norm and an unobserved social norm that affects tax compliance and employment behavior.

A series of municipal control variables are included. The sectoral composition of the municipal economy is controlled for using the shares of agriculture, manufacturing, services, and public sector production in the total municipal gross product. While factors that affect informality may influence the structure of the local economy in the long run, the concern here is that the relative demand for informal

labor (and products and services from the informal economy) may be higher in local economies where certain sectors dominate (such as agriculture or services). An explanation by Bosch et al. (2006) for why informality has increased in Brazil is that it is partly "due to a normal reallocation of workers to a sector that is intrinsically informal" (p. 25), by which they refer to production of non-traded goods, such as local services. Average firm size (number of employees) in the formal sector is included to control for technology in the formal sector, under the assumption that labor is more productive in large firms, which would increase the expected income in the formal sector for any level of worker-skill. A distance-weighted measure of local population size is included to control for market potential, assuming that the higher the market potential the higher the returns to acting formally. The share of rural-to-urban migrants in the labor force is added under the assumption that it will increase the relative supply of unskilled labor and drive down the labor income in the informal sector. Lastly, municipal product per capita is included to check to what extent the results remain robust while keeping income level constant. Since average income is most likely determined by several of the other municipal variables in the model (including government effectiveness), the results of the models that control for income should be interpreted with caution. Additional variables are used in the instrumental-variable approach and are discussed separately in the next section.

6. EMPIRICAL RESULTS

6.1. Individual-level probit model

Table 3 contains the results of the probit model specified in equations (13)–(15). The binary variable *is*, which indicates whether the worker is employed in the informal sector (1) or not (0), serves as the dependent variable in all specifications. To be consistent with the notation above, variables are grouped into human capital (s), other individual characteristics (x), municipal key variables (r), and municipal control variables (z). Six specifications are reported, in which variables have been added stepwise to evaluate the validity of the hypotheses and assess the stability of the coefficient estimates as additional factors are controlled for. The coefficients show marginal effects of small changes in the independent variables or changes from zero to one for binary variables. Standard errors are adjusted to take into account clustering of the error term within municipalities. Failure to account for intra-group correlation of the error term in multi-level data analysis may lead to rejection of the null-hypothesis of a zero-value coefficient far too easily (Primo et al., 2007).[3] Due to the large sample size (more than 2 million observations), standard errors are still small after this adjustment and most of the coefficients are statistically significant at very high levels. Specific attention should therefore be given to the economic significance when interpreting the coefficient estimates.

      First, in line with the model prediction, human capital has a negative effect on the probability of working in the informal sector. Age has a negative but decreasing effect, while years of education appears to have a negative and increasing effect, as indicated by the coefficient estimates of the square terms of these variables. This relationship remains stable across the specifications, as municipal-level variables are added. Second, both of the coefficients of the key municipal variables of interest – *government effectiveness* (g) and the strength of the *social norm* (n) – are of expected sign and in most cases statistically significant.

---

[3] Regression results without this standard error adjustment, not reported here, indicated significance at the 1-percent level for practically all municipal-level coefficients.

Column 3 of Table 3 shows a coefficient of government effectiveness of about –0.18. When controlling for other municipal factors, the coefficient decreases to about –0.13. This means that an increase of the index value of g by one standard deviation, from the average of 0.35 to 0.49, while holding everything else constant, is associated with a two-percentage point decline in the probability of a worker being informally employed. A similar exercise for the norm coefficient suggests that a one standard deviation increase in the strength of the social norm in the municipality decreases the probability of informality by a magnitude of about six percentage points.

To gain further insight into how government effectiveness might affect informality, the g index is decomposed into its three sub-components and included in the probit model. The results in column 5 suggest that resources for policy formulation (g_1) are less important than the quality of bureaucracy (g_2) and public service provision (g_3). One interpretation of these results relates to the patterns of public trust in politicians, as discussed in the beginning of the paper; political promises delivered by means of planning, regulation, and the creation of municipal ordinances may have little effect on citizens' incentives if they doubt that these political efforts will have any real effect on them (Saavedra and Tommasi, 2007). Quality of bureaucracy and public services, on the other hand, might have a more direct effect on incentives in terms of actual enforcement of regulation and other value-added in the formal sector.

The only case in which the government effectiveness coefficient becomes statistically insignificant is when municipal gross product per capita is included as a control variable (column 6). While the social norm coefficient remains significant, this suggests that the level of economic development has a more important effect on informality than government effectiveness. However, income per capita should probably be perceived, to a larger extent than any other control variable, as endogenously determined by other municipal characteristics. This potential endogeneity calls for caution in the interpretation of the results when income is included (this is discussed further in Section 6.3.).

A few observations are warranted regarding the coefficients in Table 3 that are not part of the core hypotheses. Ten individual characteristics other than human capital proxies are included in the analysis. First, being female has a very small positive effect on the probability of being in the informal sector. This gender effect is notable only where there are young children in the household (according to the interaction coefficient). Disability slightly reduces the chances of participating in the formal sector, and so does being a recent migrant from a rural area. The specification in column 2 contains two additional household characteristics: earned income among other household members and an indicator showing if any other household member is employed in the formal sector. The positive effect of household income suggests that one's incentives to take a job in the formal sector are weakened if one can be supported by other household income. The negative coefficient of formal-sector employment of another household member, however, suggests the opposite: if a household member works in the formal sector, the individual is much more likely to do so too. The correlation of sector participation among household members is quite high (0.45). This high correlation, together with the large rise in pseudo-R^2 when this variable is included, raises some concerns about endogeneity. Although the inclusion of these household variables does not affect the other coefficients to a large extent, they are not included in any of the other specifications reported in Table 3.

Among the coefficients of the local economy control variables (z), it can be noted that the share of manufacturing in the local economy has a negative effect on informality. The share of agriculture has a negative effect as well, even though weaker. The relative size of the service sector shows no significant relationship with informality (the left-out sector is the public sector). There is a positive relationship

between informality and the share of the urban labor force that consists of recent rural-to-urban migrants (even when controlling for migrant status of the individual). The presence of migrants from rural areas is likely to push down earnings in the informal sector, which would increase the incentives to look for jobs in the formal sector. It could be, though, that the probability of finding a job in the formal sector is also affected negatively by the presence of migrants, and that this effect outweighs the increased incentive due to earnings differentials. Lastly, the size of the local population has a small negative effect on informality, giving some support to the notion that there are higher returns to formality in larger markets. This could be due to greater opportunities for specialization as well as economies of scale in production.

6.2. Municipal-level fractional logit model

The municipal-level fractional logit model serves as a complement to the worker-level probit model. Besides constituting a robustness check of the municipal-level findings in the probit model, it allows for an alternative interpretation of the relationship between informality and the explanatory municipal variables. The dependent variable in this model is the share of the municipal urban labor force that is employed informally. The key variables of interest are the same as in the previous model, with the difference that individual human capital is aggregated here to the municipal average. Table 4 gives an overview of the results and includes five model specifications. Similar to the probit analysis, the composite governance index g is used in the first specifications, and is then disaggregated into its sub-indices. The coefficients show marginal effects of a small change in each explanatory variable. The results are qualitatively, and in terms of statistical significance, similar to the probit model. The coefficients are not directly comparable to the probit model, since this model contains a fractional, and not a binary, dependent variable.

The coefficient for average education is around –0.08 in the different specifications, and the interpretation is straightforward: holding everything else constant, a one-year increase in average education is associated with an eight-percentage-point decrease in the share of informal labor employment in the average municipality. The government effectiveness coefficient is about -0.16 when no municipal control variables are included (column 1). It is reduced to about half that magnitude once other municipal characteristics are accounted for. Since the index lies between zero and one, the coefficient may be interpreted as an elasticity: a ten-percentage-point increase in the g index is associated with a moderate decline of informality of about one percentage point. The same interpretation applies for the social norm, which has a coefficient between –0.4 and –0.2. A ten-percentage-point increase in the norm index is associated with a two to four-percentage-point decline in informality, when other factors are held constant. In both models, the social norm coefficient is consistently larger than the governance coefficient. This gives some support to the notion that *informal* institutions (understood here as socially sanctioned norms of behavior) play a stronger role in shaping the employment outcome and willingness to participate in the formal economy than the local government's enforcement and implementation of the *formal* institutions.

The disaggregation of the governance index g in column 4 yields a slight deviation from the results of the probit model. While all three coefficients are still negative, policy formulation (g_1) has a larger coefficient than public goods provision. Contrary to the probit results, this suggests that policy formulation matters as well.

The inclusion of income per capita (column 5) has the same effect as in the probit model; the coefficient of the governance index is rendered insignificant, while the social norm coefficient is smaller but still significant. The same endogeneity concerns apply here as in the probit model. Among the effects of the local economy, it can be seen that the shares of manufacturing and services in the local economy have elasticities ranging from −0.1 to −0.2 and 0.1 to 0.2, respectively. The municipal-level analysis confirms the role of the size of the local economy, represented by the population measure.

6.3. Robustness of the results

The robustness checks of the results are concerned primarily with the potential endogeneity of several of the explanatory variables. The first and main suspect for endogeneity is the government effectiveness index (g). It is plausible that the quality of governance is affected by economic and human development together with a range of unobserved factors, which also affect the level of informal employment in a region. An instrumental-variable approach is used in order to take into account this potential endogeneity. Instrumental variable candidates are needed that are correlated with g, yet uncorrelated with the error terms of the two models. Three sets of instrumental variables are used for this purpose, which are largely inspired by other studies. These are summarized as *geography*, *ethnicity*, and *political history* and are discussed in turn below.

Regarding the relationship between geography and institutional quality, Naritomi et al. (2007), who use governance as an indicator of institutional development in regions of Brazil, find that a set of geographical variables are significantly related to both economic and institutional development. Almeida and Carneiro (2009) analyze the effect of labor regulation enforcement on informal employment. They suspect that enforcement may be endogenous and use distance to the nearest enforcement office interacted with the local intensity of labor inspectors as an instrumental variable. Nee and Opper (2009) suggest that the quality of bureaucracy may be affected either positively or negatively by the size of a country, in terms of area. While smaller countries tend to be more homogeneous and may respond more easily to citizens' preferences, large countries might benefit from economies of scale in bureaucracy. Inspired by these studies, four variables are included as geographical instruments: latitude, longitude, area of the municipality, and transportation cost to the state capital.

The second set of instruments is inspired by La Porta et al. (1999). They find some empirical evidence that governance performance is worse in countries with higher ethno-linguistic fractionalization and where Catholicism or Islam is the dominant religion. Ethno-linguistic fractionalization is not a pronounced problem in Brazil. To capture another aspect of possible ethnic fractionalization, *racial* fractionalization is used as an instrument. Moreover, the share of Roman Catholics in the municipal population is used to capture the possible effect of religion on governance.

The age of the municipality represents the third category of instruments. More than 1 400 of Brazil's over 5 500 municipalities were created after the constitutional reform in 1988. There is some anecdotal evidence that some of these municipalities were partly created out of rent-seeking motives and that governance performance has developed quite poorly in some of these new municipalities (IBGE, 2001; The Economist, 2008). When g is regressed on the instruments and the other exogenous municipal variables, all the instruments, except racial fractionalization, show a statistically significant conditional correlation with g. The instrumental-variable approach is applied both to the probit model and to the municipal-level model.

For the probit model, specification (4) from Table 3 is used as a benchmark, which includes the composite g index. Table 5 shows the results of the analysis. The author has been unable to find a reliable method to recover the marginal effects from an instrumental probit model; therefore the raw coefficients are shown in column 1. Converting the coefficients in column 1 to marginal effects reproduces column 4 in Table 3. Column 2 of Table 5 reveals that the geographical and political history instruments are statistically significant but the ethnicity instruments are not. The coefficients of the worker characteristics are just minimally affected by the IV approach compared to the regular probit model. Considering the municipal key variables, the g index gets a larger negative coefficient as a result of using instrumental variables, whereas the coefficient on the social norm gets a smaller negative coefficient. The coefficients of the local economy control variables are altered in magnitude but not in their signs.

For the municipal-level model, the method of two-stage least squares is used, which is compared to ordinary least-squares results. Table 6 presents the results of the instrumental approach as follows: Two model specifications are used as benchmarks. The first uses the aggregated g index and corresponds to model (3) in Table 4. Columns 1–3 in Table 6 have this specification. The second specification uses the disaggregated g_1, g_2, and g_3 indices and corresponds to model (4) in Table 4.[4] Columns 4–6 in Table 6 have this specification. A comparison of columns 1 and 2 shows that the results of the fractional logit ("flogit") are in fact very similar to the results of the linear (OLS) probability model. The instrumental specification reported in column 3 shows results that are qualitatively similar to the OLS specification, but the magnitudes of the coefficients change. In absolute magnitudes the coefficient of labor force skill declines, and the coefficients for social norm and the instrumented governance variable increase. The IV results of the specification that uses the disaggregated g indices are somewhat inconclusive; in column 6 informality is positively related to public goods, while the negative relationships with the two other governance indicators remain. In addition, the coefficient of the social norm becomes statistically insignificant.

In sum, the instrumental-variable approach for both the individual-level model and the municipal-level model confirms the results in Tables 3 and 4, but when the governance index is disaggregated so that all three sub-indices are instrumented for, some noticeable deviations are observed. The choice of instruments plays a role when all three indices are instrumented for. Some further endogeneity concerns and limitations of the empirical results are elaborated on briefly in the remainder of this sub-section. The regression output of the six additional robustness checks, which are discussed here but not included in the paper, are available from the author upon request.

First, per-capita income, as a general indicator of local economic development, is likely to affect – and be affected by – several observable and unobservable variables (including quality of governance). Due to this potential endogeneity, it is included only in one specification of each model. This resulted in a statistically insignificant coefficient estimate of the governance index. To assess the importance of the municipal key variables while still holding local per-capita income "constant", the models were evaluated on a sub-sample of municipalities with relatively homogenous income. A 40-percent sub-sample of municipalities, consisting of the 2 113 "middle-income" municipalities with per-capita income of between 2 000 and 5 000 R\$, was selected for this purpose. While coefficient estimates change slightly in magnitude, no qualitative changes occur with this sub-sample. A tentative conclusion is that the results are not driven to any large extent by differences in productivity or per-capita income.

---

[4] Attempts were made to instrument the disaggregated indices g_1, g_2, and g_3 in the probit model, but without success of finding concavity of the likelihood function. Therefore IV results in which g_1, g_2, and g_3 are separated are only available for the municipal-level.

Second, the structure of the local economy could be endogenously determined in the model, just as governance might be. The structure of the local economy could be affected by human capital intensity, institutional and economic development, geography, or by unobserved characteristics. The models were therefore evaluated on sub-samples with relatively homogenous structures of the local economy. While municipalities dominated by services and manufacturing do not deviate from the previously obtained results, agriculture-dominated municipalities do not show the same strong relationships between informality and governance.

Third, the empirical literature on spatial human capital externalities is usually concerned about endogenous sorting of skilled people to certain regions (Moretti, 2004). Thus the average level of human capital in a city or region might not be exogenously determined. While it is outside the scope of this study to fully satisfactorily adjust for this possible endogeneity (by the means of additional instrumental variables), the method of evaluating the models on sub-samples has been applied here as well. By using the sub-sample of individuals who have never moved from one municipality to another, some of the endogenous-sorting problem is taken care of. Even if the resulting sample size decreases by half, the key coefficient estimates remain robust, with only minor changes in magnitude. Another sub-sample includes only municipalities with relatively homogenous education (those less than ½ standard deviation away from the average level of education). Parameter estimates from regressions on this sample are similar to those on the full sample, with the exception for two of the disaggregated governance indices in some of the specifications.

Fourth, in the results discussed above there is no distinction made between the self-employed and employees in the formal sector. While these two groups are treated as homogenous in this study, they might face different income prospects and hence different incentives regarding sector choice. Some empirical studies focus solely on employees (Pratap and Quintin, 2006) or self-employed (Blau, 1985) in the informal sector, due to their potentially fundamental differences. No major deviations in the results are observed when the probit model is estimated on sub-samples with a) all informal employees excluded and b) with all informal self-employed excluded.

Fifth, there is some regional variation in the results. When the models are estimated on each of the five macro regions in Brazil, statistically significant relationships cannot be established between informal employment and the disaggregated governance indices and the social norm indicator for the North or the Center-West. These two regions are sparsely populated and together only account for 14 percent of the sample.

Last, the models have been tested for sensitivity to certain outliers. The models have been estimated with the "tail" municipalities excluded, defined here as those that are at least two standard deviations away from the average level of informality and education. The results do not change noticeably. In sum, few of the results of the sub-sample evaluations of the models deviate from the findings in Tables 3 and 4. The cases in which key coefficient estimates do come out insignificant, or even of the opposite sign, are those in which small sub-samples are used, such as the Northern and Center-West macro regions, and agriculture-dominated municipalities.

The conclusion from these additional robustness checks is the same as above; the coefficient estimates of main interest generally remain stable and significant with the expected sign. In cases where the sample size is shrunk to a small subset, some deviation in the results is observed.

7. CONCLUSION

This paper aims at shedding some new light on the question of what causes informal employment to vary across regions. It complements the existing empirical literature by seeking to fill the gap between micro-level studies – which explain informal employment as something determined by individual characteristics – and cross-country studies – which explain it in terms of differences in tax systems, labor regulation, and quality of institutions and governance. The challenge in this case is to explain the within-country differences in informality observed in Brazil, ranging from 20 percent of the urban labor force in some municipalities (which is comparable to the informality in Chile) to 80 percent in others (which is comparable to Paraguay).

A theoretical model is proposed to explain these differences in terms of worker skills, local quality of governance, tax rates, enforcement of tax and labor regulations, and non-pecuniary costs and benefits in the informal sector. The empirical assessment of the model supports the main hypotheses: informality is higher where education is lower, where governance is less effective, and where social norms on tax compliance are weaker. There is also some evidence that social norms have a stronger effect on employment outcome than the authorities' enforcement and implementation of formal institutions. These results complement previous studies by showing that regional factors, which are exogenous to the individual, affect individual employment outcomes. Moreover, they complement the cross-country studies by showing that regional differences within a country may cause informality to vary just as much as between countries, despite the fact that the legal system, labor and tax regulation, and other formal institutions all are held constant.

The empirical strategy relies on several indices to represent government effectiveness. Although results are presented for disaggregated indices, to separate the effects of bureaucracy resources, public service provision etc., the question that ultimately remains is exactly what kind of government action is most effective in terms of including its citizens in the formal economy. The analysis in this paper is not specific enough to provide the answer. To prevent "exit and exclusion" from the formal economy (to use the title of a recent World Bank report on the topic; Perry et al., 2007), the results suggest that the worker's incentive structure needs to be taken into consideration. Without strong incentives to engage in the formal economy, the worker will either "opt out" from it or avoid trying to engage in it in the first place. While education will increase the worker's chances to overcome the skill threshold to the formal sector, flexible labor legislation could probably improve the prospects for some workers to find a job in the formal sector.

The incentive structure need not only contain economic aspects. For the local government the most challenging task – besides providing education, efficient bureaucracy, and other public services to its citizens – might be to improve the quality of the "social contract" between the authorities and the citizen, as discussed by Saavedra and Tommasi (2007). This implies making participation in the formal sector the norm rather than an exception. While such norms are likely to change only slowly over time, a government can seek ways to improve the sense of political participation and inclusion among its citizens. This includes transparency in the political decision process and in the spending of public resources, as well as recognizing the needs of the people outside the formal sector just as much as the needs of those who are already in it.

REFERENCES

Almeida, R. & Carneiro, F. (2009). Enforcement of Labor Regulation and Firm Size. *Journal of Comparative Economics, 37* (1), 28–46.

Amuedo-Dorantes, C. (2004). Determinants and Poverty Implications of Informal Sector Work in Chile. *Economic Development and Cultural Change*, *52* (2), 347–368.

Blau, D. (1985). Self-Employment and Self-Selection in Developing Country Labor Markets. *Southern Economic Journal*, *52* (2), 351–363.

Bosch, M., Goni, E., & Maloney, W. (2007). The Determinants of Rising Informality in Brazil: Evidence from Gross Worker Flows. World Bank Policy Research Working Paper 4375.

Chaudhuri, K., Schneider, F., & Chattopadhyay, S. (2006). The Size and Development of the Shadow Economy: An empirical Investigation from States of India. *Journal of Development Economics, 80* (2), 428–443.

de Paula, A. & Scheinkman, J. (2007). The Informal Sector. NBER Working paper 13486.

Economist (2008). Why Elections, Even to Lowly Office, Are So Expensive. September 25. [http://www.economist.com/world/americas/displaystory.cfm?story_id=12305155]

El Badaoui, E., Strobl, E., & Walsh, F. (2008). Is There an Informal Employment Wage Penalty? Evidence from South Africa. *Economic Development and Cultural Change*, *56* (3), 683–710.

Evans, P. & Rauch, J. E. (1999). Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth. *American Sociological Review*, *64* (5), 748–765.

Fields, G. S. (1975). Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs. *Journal of Development Economics, 2* (2), 165–187.

Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobatón, P. (2000). Dodging the Grabbing Hand: the Determinants of Unofficial Activity in 69 Countries. *Journal of Public Economics*, *76* (3), 459–493.

Funkhouser, E. (1996). The Urban Informal Sector in Central America: Household Survey Evidence. *World Development*, *24* (11), 1737–1751.

Galiani, S. & Weinschelbaum, F. (2007). Modeling Informality Formally: Households and Firms. CEDLAS working paper 47.

Gong, X., van Soest, A., & Villagomez, E. (2004). Mobility in the Urban Labor Market: a Panel Data Analysis for Mexico. *Economic Development and Cultural Change 53* (1)*, 1–36.

Harbaugh, W. T. (1996). If People Vote Because They Like to, Then Why Do So Many of Them Lie? *Public Choice*, *89* (1–2), 63–76.

Henley, A., Arabsheibani, G. R., & Carneiro, F. G. (2009). On Defining and Measuring the Informal Sector: Evidence from Brazil. *World Development, 37* (5), 992–1003.

IBGE (2001). IBGE traça o perfil administrativo de todos os municípios brasileiros. Press release, April 17, 2001. [http://www1.ibge.gov.br/home/presidencia/noticias/1704munic.shtm]

Johnson, S., Kaufmann, D., & Zoido-Lobatón, P. (1998). Regulatory Discretion and the Unofficial Economy. *American Economic Review, 88* (2), 387–392.

Jütting, J., Parlevliet, J., & Xenogiani, T. (2008). Informal Employment Re-loaded. Working Paper 266, OECD Development Centre.

Kucera, D. & Xenogiani, T. (2009). Women in Informal Employment: What Do We Know and What Can We Do? In J. Jütting & J. R. de Laiglesia (Eds.), *Is Informal Normal?: Towards More and Better Jobs in Developing Countries* (pp. 89–113). Paris: OECD Development Centre.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The Quality of Government. *Journal of Law, Economics and Organisations*, *15* (1), 222–279.

Loayza, N. (1996). The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy, 45*, 129-162.

Loayza, N. & Rigolini, J. (2006). Informality Trends and Cycles. World Bank Policy Research Working Paper 4078.

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

Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. (1997). Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change, 45* (2), 367–392.

Mazumdar, D. (1976). The Rural-Urban Wage Gap, Migration, and the Shadow Wage. *Oxford Economic Papers*, 28 (3), 406–425.

Moretti, E. (2004). Human Capital Externalities in Cities. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics, Volume 4* (pp. 2243–2291). Amsterdam: Elsevier Science.

Naritomi, J., Soares, R. R., & Assunção, J. J. (2007). Rent Seeking and the Unveiling of 'de facto' Institutions: Development and Colonial Heritage within Brazil. NBER Working Paper 13545.

Nee, V. & Opper, S. (2009). Bureaucracy and Financial Markets. *Kyklos, 62* (2), 293–315.

Papke, L. & Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, *11* (6), 619–632.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., & Mason, A. (2007). *Informality: Exit and Exclusion.* Washington, D.C.: World Bank.

Pisani, M. & Pagán, J. (2004). Sectoral Selection and Informality: a Nicaraguan Case Study. *Review of Development Economics*, *8* (4), 541–556.

Pradhan, M. & van Soest, A. (1995). Formal and Informal Sector Employment in Urban Areas of Bolivia. *Labour Economics, 2* (3), 275–297.

Pratap, S. & Quintin, E. (2006). Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach. *European Economic Review*, *50* (7), 1817–1841.

Primo, D. M., Jacobmeier, M. L., & Milyo, J. (2007). Estimating the Impact of State Policies and Institutions with Mixed-Level Data. *State Politics and Policy Quarterly*, *7* (4), 446–459.

Rauch, J. E. (1991). Modelling the Informal Sector Formally. *Journal of Development Economics*, *35* (1), 33–47.

Saavedra, J. & Tommasi, M. (2007). Informality, the State and the Social Contract in Latin America: A Preliminary Exploration. *International Labour Review*, *146* (3–4), 279–309.

Schneider, F. (2005). Shadow Economies around the World: What Do We Really Know? *European Journal of Political Economy*, *21* (3), 598–642.

Schneider, F. & Enste, D. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature, 38* (1), 77–114.

Straub, S. (2005). Informal Sector: The Credit Market Channel. *Journal of Development Economics*, *78* (2), 299–321.

Torgler, B. & Schneider, F. (2007). Shadow Economy, Tax Morale, Governance and Institutional Quality: A Panel Analysis. CESifo Working Paper 1923.

World Economic Forum (2008). *The Global Competitiveness Report 2008–2009*. Geneva.

TABLE 1. *Characteristics of the urban labor force, divided by formal and informal sector*

	Age (percent in each age category)					
	15-24	25-34	35-44	45-54	55-65	average age
Formal sector	21	31	27	16	5	35
Informal sector	28	27	23	15	7	34
	Years of education (percent in each category)					
	0-2	3-5	6-8	9-12	>12	average years
Formal sector	8	21	19	36	16	8.6
Informal sector	19	32	22	22	5	6.1
	Earnings, at percentile (R$ per month)					
	20%	40%	60%	80%	average	
Formal sector	221	330	500	970	769	
Informal sector	120	167	300	500	404	
	Sector of employment (percent per sector)					
	Commerce	Manufacturing	Construction	Domestic services	Other services	
Formal sector	21	13	12	5	46	
Informal sector	18	18	5	14	30	

Note: In August 2000, the exchange rate was R$1 = US$0.56. *Source*: Brazilian Demographic Census, 2000.

TABLE 2. *Variables used in the empirical analysis*

Variable	Mean	S.D.	Definition
			Individual level (N = 2,222,387)
Employment and human capital			
Informal employment	0.436	0.496	Individual works in informal sector, with earned non-zero income
Age	34.5	11.6	Individual's years of age
Education	7.5	4.3	Individual's years of schooling
Other individual characteristics (x)			
Gender - female	0.397	0.489	Individual is female
Household head	0.516	0.500	Individual is household head
Married	0.454	0.498	Individual is married
Race - black	0.066	0.248	Individual is black
Disabled	0.021	0.144	Individual has reduced eyesight, hearing, paralysis, or mental problem
Rural-urban migrant	0.029	0.169	Individual has moved from a rural area 5 years or less, prior to the survey
Children in hhd	0.480	0.500	There is at least one child in the household, 10 years old or younger
Formal-sector worker in hhd	0.311	0.463	There is at least one other household member employed in the formal sector
Household income	859	2,887	Monthly total income of other household members
			Municipal level (N = 5,506)
Share informal employment	0.554	0.161	Share of the urban municipal labor force that has informal employment.
Average education in labor force	6.4	1.1	Average years of education among workers in the municipal labor force.
Municipal key variables (r)			
Government effectiveness, g	0.348	0.139	Index composed of g_1, g_2, and g_3, below, to represent government effectiveness in the municipality.
Policy formulation, g_1	0.253	0.183	Index composed of 20 variables to represent the capacity of the municipal government to formulate and implement policy.
Bureaucratic resources, g_2	0.427	0.131	Index that represents the human, technical, and managerial resources available to the municipal bureaucracy.
Public goods, g_3	0.363	0.234	Index composed of 10 variables to represent the quality of, and access to, public goods in the municipality.
Social norm	0.835	0.063	Index representing the strength of the social norm in the municipality that affects the moral cost of acting informally. Defined as the average share of the eligible voters who voted in the municipal and presidential elections, year 2000 and 2002, respectively.

TABLE 2. *(Continued)*

Variable	Mean	S.D.	Definition
Local economy (z)			
Share agriculture	0.280	0.183	Agriculture as share of gross municipal product.
Share manufacturing	0.190	0.162	Manufacturing as share of gross municipal product.
Share services	0.504	0.158	Services as share of gross municipal product.
Share public sector production	0.270	0.141	Public sector production share of gross municipal product.
Average firm size	5.4	5.7	Average number of employees in registered businesses in the municipality.
Local population	694	1,464	Population in surrounding municipalities, weighted by distance (thousands)
Gross product per capita	4,435	5,699	Municipal gross product per capita, R$, year 2000.
Share rural immigrants	0.056	0.047	Share of the urban labor force consisting of workers who have migrated from a rural area, five years or less prior to the Census survey year.

Sources: Individual-level variables and informal employment and labor force skill on municipal level – Brazilian Demographic Census, 2000. Municipal variables for *government effectiveness, social norm, and local economy* – Gestão Pública, 2005–2006, and Base de Informações Municipais, 2000.

TABLE 3. *Estimation results – probit model (probability of informal employment)*

	(1)	(2)	(3)	(4)	(5)	(6)
Human capital (s)						
Age	-0.023***	-0.027***	-0.022***	-0.022***	-0.022***	-0.022***
Age, squared	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Education	-0.029***	-0.022***	-0.023***	-0.022***	-0.022***	-0.021***
Education, squared	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
Other individual char's (x)						
Gender - female	0.007**	-0.001	0.008**	0.008**	0.008**	0.007**
Household head	-0.031***	-0.057***	-0.033***	-0.033***	-0.033***	-0.034***
Married	-0.052***	-0.025***	-0.051***	-0.049***	-0.049***	-0.049***
Race - black	-0.012***	-0.006**	-0.011***	-0.010***	-0.010***	-0.009***
Disabled	0.052***	0.037***	0.048***	0.047***	0.047***	0.046***
Rural-urban migrant	0.030***	0.035***	0.019***	0.011***	0.011***	0.014***
Children in household	0.018***	0.005***	0.010***	0.009***	0.009***	0.007***
Female X chd in hhd	0.047***	0.045***	0.048***	0.047***	0.047***	0.048***
Household income		0.035***				
Formal-sector worker in hhd		-0.521***				
Municipal key variables (r)						
Government effectiveness, g			-0.178***	-0.129***		-0.014
Policy formulation, g_1					-0.030	
Bureaucratic resources, g_2					-0.065***	
Public goods, g_3					-0.045***	
Social norm, n			-1.247***	-0.963***	-0.952***	-0.668***
Local economy (z)						
Gross product per capita						-0.071***
Share agriculture				-0.112**	-0.109**	-0.056
Share manufacturing				-0.233***	-0.234***	-0.157***
Share services				-0.025	-0.024	-0.101*
Average firm size				-0.001	-0.001	0.001
Share rural immigrants				0.210**	0.205**	0.213**
Local population				-0.012***	-0.012***	-0.008***
Sample size	2,220,387	2,220,387	2,218,167	2,213,429	2,213,429	2,213,429
McFadden pseudo R-squared	0.070	0.240	0.085	0.088	0.088	0.091

Note: Dependent variable is the dummy variable indicating informal employment (*is*). Coefficients show marginal effects. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for error term clustering. Coefficient estimates that are reported as "0.000", and yet statistically significant, are smaller than 0.0005 in absolute magnitude.

TABLE 4. *Estimation results – municipal fractional logit (share informal employment).*

	(1)	(2)	(3)	(4)	(5)
Labor force skill					
Average education in labor force	-0.086***	-0.080***	-0.077***	-0.077***	-0.058***
Municipal key variables					
Government effectiveness, g	-0.160***	-0.077***	-0.071***		0.013
Policy formulation, g_1				-0.027***	
Bureaucratic resources, g_2				-0.037***	
Public goods, g_3				-0.016**	
Social norm, n	-0.425***	-0.314***	-0.212***	-0.209***	-0.073***
Local economy					
Share agriculture		-0.031	-0.014	-0.017	0.032
Share manufacturing		-0.189***	-0.123***	-0.127***	-0.009
Share services		0.138***	0.172***	0.168***	-0.074**
Average firm size		-0.002***	-0.001***	-0.001***	-0.001***
Share rural immigrants		0.058	-0.03	-0.029	0.015
Local population			-0.024***	-0.024***	-0.021***
Gross product per capita					-0.107***
Sample size	5,500	5,458	5,434	5,434	5,434
Akaike Information Criterion	0.896	0.893	0.890	0.891	0.885

Note: Dependent variable is the share of the municipal labor force in the informal sector (*IS*). Coefficients show marginal effects. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity.

TABLE 5. *Probit model with instrumental variables (probability of informal employment)*

	(1) PROBIT	(2) IV
Human capital		
Age	-0.056***	-0.056***
Age, squared	0.001***	0.001***
Education	-0.057***	-0.055***
Education, squared	-0.001***	-0.001***
Municipal key variables		
Government effectiveness, g	-0.331***	-0.928**
Social norm, n	-2.469***	-2.233***
Instrumental variables		
Transp. cost to state capital		-0.000**
Municipal Area		0.000***
Latitude		-0.007***
Longitude		-0.005***
Racial fractionalization		0.11
Share Roman-Catholic		0.018
Age of municipality		0.000***
Sample size	2,213,429	2,213,429

Note: Dependent variable is the dummy variable indicating informal employment (*is*). Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for error term clustering. Coefficient estimates that are reported as "0.000", and yet statistically significant, are smaller than 0.0005 in absolute magnitude.

TABLE 6. *Municipal model with OLS and instrumental variables (share informal employment)*

	(1) FLOGIT	(2) OLS	(3) IV	(4) FLOGIT	(5) OLS	(6) IV
Labor force skill						
Average educ. in labor force	-0.077***	-0.073***	-0.041***	-0.077***	-0.073***	-0.035***
Municipal key variables						
Government effectiveness, g	-0.071***	-0.071***	-0.756***			
Policy formulation, g_1				-0.027***	-0.028***	-1.124***
Bureaucratic resources, g_2				-0.037***	-0.034***	-0.998***
Public goods, g_3				-0.016**	-0.017**	0.916***
Social norm	-0.212***	-0.209***	-0.310***	-0.209***	-0.206***	-0.031
Local economy						
Share agriculture	-0.014	-0.014	-0.323***	-0.017	-0.016	-0.593***
Share manufacturing	-0.123***	-0.122***	-0.312***	-0.127***	-0.125***	-0.728***
Share services	0.172***	0.164***	-0.176***	0.168***	0.160***	-0.563***
Average firm size	-0.001***	-0.001***	-0.002***	-0.001***	-0.001***	-0.003***
Share rural immigrants	-0.03	-0.031	-0.037	-0.029	-0.029	0.202**
Local population	-0.024***	-0.023***	-0.020***	-0.024***	-0.023***	-0.026***
Constant		1.300***	1.703***		1.304***	1.925***
Sample size	5,434	5,434	5,434	5,434	5,434	5,434
R^2		0.556			0.556	

Note: Dependent variable is the share of the municipal labor force in the informal sector (*IS*). FLOGIT refers to fractional logit, OLS to ordinary least squares, and IV to 2-stage least squares. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. FLOGIT columns are identical to columns 3 and 4 in Table 4. Instrumental variables are the same as in Table 5.

FIGURE 1. *The skill threshold that equates expected earnings in the formal and informal sector*
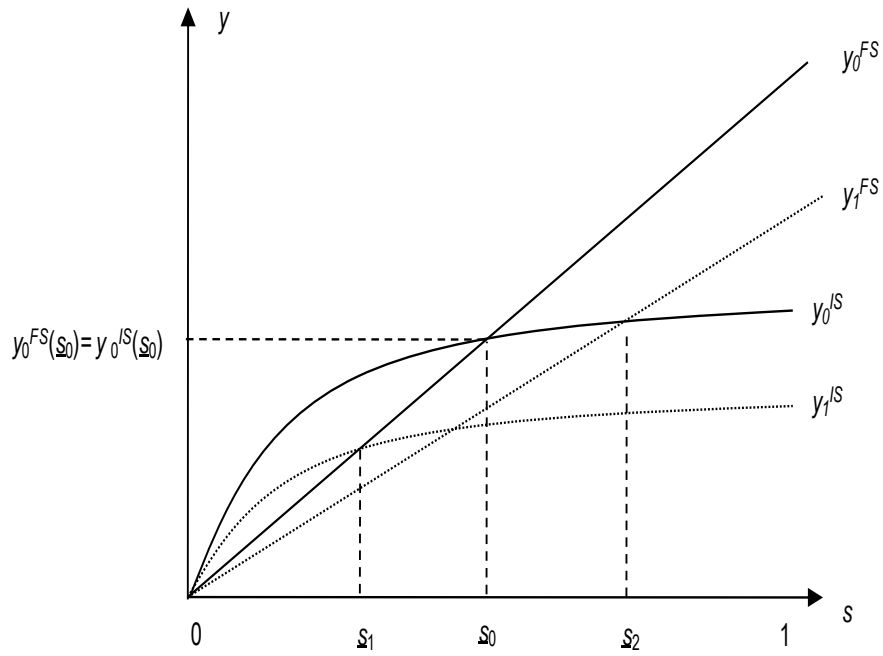

FIGURE 2. *Skill thresholds in the presence of non-pecuniary effects in the informal sector*