DISCUSSION PAPER SERIES

No. 5545

INEQUALITY AND INFORMALITY

Alberto Chong and Mark Gradstein

PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5545.asp**

ISSN 0265-8003

INEQUALITY AND INFORMALITY

Alberto Chong, Inter-American Development Bank
Mark Gradstein, Ben Gurion University of the Negev and CEPR

Discussion Paper No. 5545
March 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5545

March 2006

ABSTRACT

Inequality and Informality*

This paper presents theory and evidence on the determinants of the size of the informal sector. We propose a simple theoretical model in which it is positively related to income inequality, more so under weak institutions, and is negatively related to the economy's wealth. These predictions are then empirically validated using different proxies of the size of the informal sector, income inequality, and institutional quality. The results are shown to be robust with respect to a variety of econometric specifications. We also find that government interventions through taxes and regulations lose much of their robustness in the presence of the above factors.

JEL Classification: D70, O15 and O17
Keywords: inequality, informal sector, institutional quality and shadow economy

Alberto Chong
Research Department
Inter-American Development Bank
Stop W-0436
1300 New York Avenue NW
Washington DC 20577
USA
Tel: (1 202) 623 1536
Fax: (1 202) 623 2481
Email: albertoch@iadb.org

Mark Gradstein
Department of Economics
Ben Gurion University
Beer Sheva 84105
ISRAEL
Tel: (972) 864 72288
Fax: (972) 864 72941
Email: grade@bgumail.bgu.ac.il

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=152471

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=129935

* Our thanks to Cesar Calderon, Norman Loayza, and Ugo Panizza for useful comments on a previous draft and to Cecilia Calderon, Gianmarco Leon, and Rudy Loo-Kung for their excellent research assistance. The findings and interpretations are those of the authors and do not necessarily represent the views of the Inter-American Development Bank or their corresponding executive directors. The standard disclaimer applies.

Submitted 05 February 2006

1. INTRODUCTION

A significant part of economic activity in developing as well as in developed countries is conducted in the informal sector. Estimates suggest that, depending on the measure used, it generates between 10 and 20 percent of the aggregate output in developed countries and more than a third in developing countries, reaching in some more than 50 percent (see Table 1 below). Concerns have been expressed with regard to the effect of informality on economic growth,[1] as well as to its impact on the erosion of the tax base with ensuing detrimental effects on publicly provided goods and services. Consequently, some recent work has studied the determinants of informality. In particular, the efforts have focused on various government interventions, such as through high tax burden (e.g., Cebula, 1997; Giles and Tedds, 2002), excessive regulation, especially of labor (as in, for example, Johnson et al., 1998; Friedman et al., 2000), and institutional quality (Friedman et al., 2000)[2].

In this paper, we take a fresh look at the determinants of the size of the informal sector. In the framework below, three main factors play a role in this regard, institutional quality, overall economic development, and income inequality. Specifically, we submit the claim that institutional quality as well as rising incomes, have a negative effect on informality, so that we expect a more developed economy to have a smaller informal sector. More interesting and novel perhaps, we argue that income inequality, in conjunction with institutional quality, is a particularly significant determinant of informality. The reason for this is that, when institutional quality is low, protection of

[1] This has been the subject of some scrutiny recently. For instance, while it has been suggested that a large informal sector implies, *inter alia*, slower economic growth (Loayza, 1996), Sarte (2000) offers a more nuanced view.

[2] Schneider and Enste (2000), Schneider and Klingmair (2003), and Schneider (2005) contain excellent surveys of this work as well as a detailed account of estimation issues.

property rights in the formal sector is weak, and resources are to a large extent up for grabs. Poor individuals whose endowments are relatively limited are at a disadvantage in extracting a larger share of the resources, hence find it beneficial to move into the informal sector, where although less productive, they are able to fully retain their production output. High inequality, exacerbated by low institutional quality, magnifies this effect implying a positive relationship between inequality and the size of the informal sector.

A simple model with these properties is first exhibited and then empirically tested in this paper using recent estimates on the size of the informal sector, different proxies, as well as different econometric approaches[3]. Overall, we find that income inequality, particularly in conjunction with institutional quality, is a statistically significant and robust determinant of the relative size of the informal sector. For example, depending on the specification, an increase in inequality level from the Mexican (a Gini of .49) to the Brazilian one (a Gini of .57) increases this size by 3 to 9 percentage points. In addition, while uncovering the salience of income inequality and institutional quality, we also include some of the government policy variables traditionally viewed as important determinants of the informal sector. Interestingly, in the presence of institutional and inequality variables, both the proxies for the tax burden and the labor rigidity turn out to be not robust, significant in some specifications but not in others. While in agreement with Friedman et al., (2000), as to the potential importance of institutional quality, we go further arguing, both in the theoretical framework and in the empirical estimation, that income inequality is a crucial factor in determining the scope of informality, in particular,

[3] From a theoretical perspective, this work is related to a number of recent papers that generate informality in equilibrium, such as Acemoglu (1995), Acemoglu and Verdier (2000), and Loayza (1996) among others. However, none of these papers focuses on income inequality.

as it interacts with institutional quality.

The next section exhibits the theoretical model, whose analysis and empirical implications are presented in Section 3; Section 4 introduces the empirical approach and the data employed; Section 5 reports the main findings of the empirical analysis, Section 6 contains some robustness checks, and Section 7 briefly concludes.

2. FRAMEWORK

Consider a two-period economy populated by a measure one of individuals indexed by i. The initial level, in period 1, of individual i's income is exogenously given at y_i, and the period 2's income level, z_i is endogenously determined. We let H denote the cdf of the initial income distribution and assume that its support is $[0, \infty)$. Production takes place in the formal sector (*FS*) and in the informal sector (*IS*). The aggregate productivity in the former is larger, but the individual access to productive technology is limited - through licensing, regulation etc. The individuals expend resources therefore, to gain access to these technologies. In contrast, in the informal sector technology is less productive, but is readily accessible by all households.

The individuals allocate resources between current consumption, c_{i1}, productive investment, k_i, and – if they operate in the formal sector – investment in rent seeking x_i, to gain access to superior production technologies, resulting in the following budget constraint:

$$y_i = c_{i1} + k_i + x_i \tag{1}$$

Imperfect credit markets are assumed, so that (1) is binding; this will imply that poor

households – being more financially constrained to invest in rent seeking - are at a disadvantage relatively to the rich when operating in the formal sector.

In the informal sector, the production function is

$$z_i = Bk_i,\ B > 0 \tag{2}$$

where B>0 is the productivity parameter in the IS.

In contrast, the aggregate productivity parameter in the formal sector, A, is larger, $A > B$. There may be several reasons for the formal sector being superior in this regard. One reason has to with better access to technological or financial opportunities. Another reason, more related to the paper's spirit, is that operating formally generates monopoly rents, especially so when institutional quality is not perfect. This then leads to our next assumption that the allocation of production opportunities among the households in the formal sector is endogenously determined through the individual rent seeking efforts as well as through the prevailing institutional strength, which determines the marginal effectiveness of individual rent seeking. Specifically, where institutions are strong marginal rent seeking efforts are insignificant in determining the allocation of production opportunities. Letting L denote the degree of institutional strength, individual i's productivity parameter is

$$a_i = A\frac{x_i^{1-L}}{\int_{i\in FS} x_i^{1-L}di} \text{ if } 0 \leq L < 1, \text{ and } a_i = A \text{ if } L = 1 \tag{3}$$

Note that the larger L is the lower is the marginal value of spending resources to gain access to the technology in the formal sector and the more equally is this access

allocated.[4]

The interpretation is that, in the informal sector, everyone has an access to some basic technology. In contrast, access to blueprints for a superior technology available in the formal sector require licensing, acquisition of which occurs through rent seeking. Assuming a simple linear technology a household's second-period income level then is

$$z_i = a_i k_i = A \frac{x_i^{1-L}}{\int_{i \in FS} x_i^{1-L} di} k_i \tag{4}$$

Because this assumption, in conjunction with the assumed credit market imperfection, is crucial for one of the main results, we pause here to discuss its implications. It is assumed that operating in the formal sector entails competition for access to technological opportunities, such as through licensing or other regulations. This competition is done by rent seeking, the agents expending resources to obtain access. Now, credit constraints imply that richer households are in a better position to successfully engage in rent seeking, and – as will be shown formally below – this relative disadvantage of the poor increases with inequality. This intuitive argument is behind our main result that inequality increases the propensity to go informal.

The individual preferences derive from current consumption c_{i1} and future consumption, which in turn, equals next-period income z_i. Assuming for simplicity symmetric logarithmic preferences, we write the expected utility:

$$V(c_{i1}, c_{i2}) = \ln(c_{i1}) + \ln(c_{i2}) = \ln(c_{i1}) + \ln(z_i) \tag{5}$$

[4] Institutional strength is assumed to be exogenously given here. In Chong and Gradstein, (2004) it is suggested that income inequality may subvert institutions.

Whereas traditional attempts to define the informal sector have focused on the firms' size and their type of technology as well as the workers' occupation, here we follow recent research that shows that businesses and workers typically base their decision to enter the formal sector on the evaluation of the relative benefits and costs of doing so (Saavedra and Chong, 1999; Maloney 1999). Therefore, the individuals first decide in which sector – IS or FS to operate. Then, in the former case, they allocate income between consumption and investment; and in the latter case – between consumption, investment and influence activities. The equilibrium consists of such mutually consistent decisions.

3. IMPLICATIONS

3.1. EQUILIBRIUM ANALYSIS

Analysis proceeds backwards, starting with the consumption-investment choices in each sector and then determining the allocation of the individuals across the two sectors.

Maximizing (5) subject to (1) and (2) for those in the informal sector we obtain (clearly, there is no investment in getting access to technologies in this case):

$$c_{i1}^{IS} = k_i^{IS} = y_i / 2, \;\; z_i^{IS} = By_i / 2 \tag{6}$$

and the utility level of

$$V_i^{IS} = \ln(y_i /2) + \ln (By_i/ 2) \tag{7}$$

Likewise, for those in the formal sector maximization of (5) subject to (1) and (4) yields:

$c_{i1}^{FS} = k_i^{FS} = y_i / (3\text{-}L)$, $x_i^{FS} = y_i (1\text{-}L)/ (3\text{-}L)$,

$$z_i^{FS} = A\,[(y_i)^{1\text{-}L}/ \int_{i\in FS} (y_i)^{1-L}\, di\,)\,(y_i / (3\text{-}L)] = A\,[(y_i^{1\text{-}L} / \int_{i\in FS} y_i^{1-L}\, di\,)\,(y_i / (3\text{-}L)] \tag{8}$$

and

$$V_i^{FS} = \ln [y_i / (3\text{-}L)] + \ln \{A [(y_i^{1\text{-}L} / \int_{i \in FS} y_i^{1-L} di) (y_i / (3\text{-}L)]\} \quad (9)$$

Anticipating the consumption-investment choices, the individuals decide in what sector to produce, by comparing the resulting expected utilities. Comparing (7) and (9), the utility differential is

$V_i^{FS} - V_i^{IS} =$

$$\ln [y_i / (3\text{-}L)] + \ln \{A (y_i^{1\text{-}L} / \int_{i \in FS} y_i^{1-L} di) (y_i / (3\text{-}L))\} - [\ln(y_i /2) + \ln (By_i/ 2)] =$$

$$\ln [2 / (3\text{-}L)] + \ln \{(A/B) (y_i^{1\text{-}L} / \int_{i \in FS} y_i^{1-L} di) (2 / (3\text{-}L))\} \quad (10)$$

Differentiation reveals that this differential increases in income. Moreover, when an individual income is small enough, (10) is negative, whereas when it is large enough (10) is positive, implying that only rich enough individuals produce in the formal sector while the poorer ones move into the informal sector. Letting y^* denote the income threshold level above which such production takes place, we write:

$$\ln [2 / (3\text{-}L)] + \ln \{(A/B) [y^{*1\text{-}L} / \int_{y>y^*} (y^{1-L} dH) (2 / (3\text{-}L)]\} = 0 \quad (11)$$

Equation (11), therefore, determines the threshold level and the relative size of the informal sector, $H(y^*)$. Moreover, this threshold is uniquely defined. To prove this point, note that differentiation reveals that the left-hand side in (11) – increases in y^*. Further, it is clearly negative when y^* is small enough and is positive when y^* is large enough, so employing a continuity argument we obtain that there is a unique y^* solving

(11).

3.2. COMPARATIVE STATICS

We now derive some comparative statics results with respect to the unique equilibrium characterized above.

Institutional quality and income levels. Differentiation reveals that the left hand side in (11) increases in L,[5] so that the better is institutional quality the smaller is the share of the informal sector. Further, suppose that the economy becomes richer. One way to capture this is by positing a new income distribution, say G(y), which dominates the previous one, H(y), in the sense of first order stochastic domination, i.e.,

$$\int_{o}^{\omega}[G(y)-H(y)]dy \leq 0 \text{ for all } \quad \omega \geq 0 \tag{12}$$

This then implies that, for any y^*, $\int_{y>y^*}(y^{1-L}dH) < \int_{y>y^*}(y^{1-L}dG)$. Since the left-hand side in (11) increases in y^*, it follows that the cutoff income level must increase.

Thus, we have proved:

Proposition 1. The size of the informal sector is negatively related to institutional quality and is positively related to the economy's level of economic development.

---

[5] Specifically, the derivative is

$$\frac{2}{3-L}-\ln y^*+\frac{\int_{y>y^*} y^{1-L}\ln y dH}{\int_{y>y^*} y^{1-L}dH}=\frac{2}{3-L}+\frac{\int_{y>y^*} y^{1-L}(\ln y-\ln y^*)dH}{\int_{y>y^*} y^{1-L}dH}>0.$$

Income inequality. To study the equilibrium effect of income inequality, suppose that the initial income distribution consists of two classes, the poor and the wealthy, whose respective relative size is P and W, with initial incomes y_P and y_W; $y_P < y_W$, for simplicity we assume that the two are equal in size, $P = W = 1/2$.[6] In general, there may be two types of equilibria: where some of the rich, in addition to all the poor, go informal and the rest of the rich operate formally; and where some of the poor, as well as all the rich become formal, and the rest of the poor operate informally. It can be easily seen, however, that the former cannot be an equilibrium: once going formal is attractive for some of the rich, it becomes attractive to do so for all the rest.[7] In equilibrium then some of the poor become informal, whereas others produce in the formal sector; let I and F denote these fractions, $I + F = P$.

After some manipulations, the equilibrium condition (11) then becomes as follows:

$$2\ln(2/(3-L)) + \ln\{(A/B)\,[y_P^{1-L} / (F\,y_P^{1-L} + W\,y_W^{1-L})]\} =$$

$$2\ln(2/(3-L)) + \ln\{(A/B)\,[1 / (F + W(y_W/y_P)^{1-L})]\} = 0 \qquad (13)$$

This condition determines F, the fraction of the poor participating in the formal sector,

[6] With an appropriate adjustment of the concept of a mean preserving spread the results go through for the more empirically valid case P > W.

[7] A formal proof is as follows. Suppose that some of the rich go informal and the rest of the rich operate formally; then the equilibrium conditions are (F, 0<F<1/2, denotes the fraction of the rich in the formal sector):

$\ln [2 / (3\text{-}L)] + \ln \{(A/B) [(y_P^{1-L} /F\, y_W^{1-L}) (2 / (3\text{-}L)]\} < 0$

and

$\ln [2 / (3\text{-}L)] + \ln \{(A/B) [(y_W^{1-L} /F\, y_W^{1-L}) (2 / (3\text{-}L)]\} =$

$\ln [2 / (3\text{-}L)] + \ln \{(A/B) [(1/F) (2 / (3\text{-}L)]\} = 0$; however, the second condition cannot possibly hold as its left-hand side is positive.

hence, $I = P\text{-}F$, the share of the population in the informal sector. Differentiation revels that this latter share increases in the income ratio between the rich and the poor y_W/ y_P.

Consider now a mean preserving spread in incomes, so that $y_W' = y_W + e$, $y_P' = y_P - e$, $e > 0$. Such spread, from (13), decreases F and increases the proportion of the population in the informal sector, I. This illustrates the existence of a positive relationship between income inequality and the size of the informal sector. Moreover, the elasticity of the income ratio y_W'/ y_P' with respect to the mean preserving spread is higher the lower the institutional quality L, implying in turn that poor institutional quality exacerbates the effect of inequality on the size of the informal sector.

Summarizing,

Proposition 2. For a two-class income distribution, the increase in income inequality causes an increase in the relative size of the informal sector. This effect is stronger the lower is institutional quality.

The intuition for the positive relationship between income inequality and informality is as follows. When inequality between the rich and the poor increases, the former invests in rent seeking more and the latter invest less. As long as the poor are present in both the formal and the informal sectors, however, they must be indifferent between the two sectors. In equilibrium, therefore, for the poor in the formal sector to have the same rate of return on capital than in the informal sector despite less investment in rent seeking, there must be less poor in the formal sector. This effect is magnified by poor institutional quality, which enhances the difference in rent seeking investments between the rich and the poor.

To sum up the empirically relevant implications, the size of the informal sector is

expected to be larger the weaker the institutions, and the larger income inequality; and the effect of the latter is magnified with weak institutions.

4. EMPIRICAL APPROACH: DATA AND SPECIFICATION

As outlined in the formal model, informality may be thought as a decision made on the basis of cost-benefit evaluations. Changes in non-wage costs, in tax and labor regulations, in workers preferences and characteristics, and changes in workers´ and firms´ productivity may change the profitability of becoming or staying in the formal sector. In this context, entering the informal sector is a decision that both firms and employees make on the basis of cost-benefit evaluations that are continuously revised and c change depending on changes in institutions, regulations, the level of economic activity and preferences. Given that there are monetary and non-monetary costs of entering the formal sector, such as fees, taxes, time regulations, and administrative procedures, benefits should be higher than those costs in order for informal sector entrepreneurs to be willing to change their legal status. When the benefits do not exist or when law enforcement is weak, many individuals and firms will find it convenient to go switch to the informal sector (Saavedra and Chong, 1999). For example, in the case of firms, the key costs of becoming or staying informal are related to the difficulty of establishing commercial relationships with formally registered firms, the difficulty in exercising property rights, and the pecuniary costs of being caught not complying with regulations. Conversely, the main costs of becoming formal are the time spent, the taxes and other pecuniary costs related to legal requirements. This is balanced by the benefits of formality, in particular, the possibility of broadening the scale of production, the

establishment of commercial relationships with bigger firms, and the access to credit in financial markets. For wage earners, on the one hand, the choice of becoming either formal or informal should be understood in terms of the choice of working in occupations with different characteristics. It is the firm that offers the job the one that decides if it is convenient to comply with the existing rules and regulations. Creating a formal job implies the inclusion of specific labor costs for the firm in terms of taxes, non-wage costs, and administrative procedures. Thus, the informal sector can be understood as comprising one group whose decision to stay as independent workers will depend on the agent's cost-benefit analysis and will not necessarily imply change in occupations, and another group where the informal sector may somehow represent an option of last resort.

Conceptually, measurement issues are probably the most challenging topic regarding studying the informal sector (Tanzi, 1999). This, not only for the obvious reasons related with uncovering a sector that, by definition, is not part of the official economy, but also, because of how the informal sector is understood. In this paper we use two empirical definitions that are consistent with the cost-benefit analysis presented in the model above.[8] In particular we use two alternative data series. The first is based on Kaufmann and Kaliberda's (1996) Macroelectric approach. According to this method the size of the informal economy is measured as a discrepancy between an indicator of the overall economic activity and the official gross domestic product. Given the high correlation between the consumption of electricity and economic activity, the growth rate of electricity consumption serves as an indicator of the evolution of the GDP[9]. Hence,

[8] The traditional widely used definition of the informal sector focuses on the size of the firm, the occupation of the workers, and the type of technology employed. Thus, an increase in the share of small businesses in total employment is interpreted as an increase in the informal sector. This approach gives contradictory measurements in relation to a cost-benefit analysis (Saavedra and Chong, 1999).

[9] This assumes that the elasticity of electricity consumption to gross domestic product should be close to

any difference between the growth of electricity consumption and the GDP growth is attributed to changes in the size of the informal economy. To calculate this measure, we use the data on total electricity consumption from the World Bank (2004). Data on real (official) gross domestic product, measured as the nominal gross domestic product deflated by the implicit gross domestic product deflator, was obtained from the International Monetary Fund (2004) using annual observations from 1990 to 2000[10]. The second, more limited in terms of observations, comes from Schneider and Klingmair (2003) and consists of recent cross-section measures, essentially, around the 1990s and are essentially based on the *demand for currency* approach.[11] The idea is that all hidden economic activity takes place using cash as the means of exchange, so that an increase in the informal economy generates an increased demand for currency. To calculate the excessive demand for money, a standard equation for currency demand is estimated along with controls which include a proxy for the tax burden[12]. Using an empirical estimation for the expected values of currency holdings, they are then re-estimated under the assumption that the tax variable takes the value of zero. The difference between these two

---

one. Kaufmann and Kaliberda (1996) perform sensitivity analysis allowing the value of the elasticity to vary across countries and time, which may account for technological changes in production process, variations of the sectoral composition of GDP and different production structures across countries. Following this, in order to account for the fact that economies may become more efficient in the use of electricity, we assume that the elasticity decreases in 0.05 from decade to decade (from 1.15 in the 60s to the value of 1 in the 90s). The results – available on request – do not significantly change when applying different elasticities or when assuming elasticity close to 1 throughout.

[10] Sources for seed values are Kaufmann et al. (1997), Loayza (1996), Lackó (1996), Giles (1999) and Schneider and Enste (2000).

[11] These data have been derived from the mimic or dymimic method (Schneider and Klingmair, 2003; Schneider, 2005). In order to convert the mimic estimation into absolute values of the informal sector one needs the currency demand approach. Thus, for the sake of simplicity we call this method the "currency demand" one. Additionally, another available data set that uses this approach is Botero et.al., (2004) who largely base their series on Schneider and Enste (2002). Schneider and Klingmair (2003) contains an updated data series of the latter. Replicating the empirical exercises using these data yields almost identical results.

[12] The basic equation is $\ln(C/M_2)_t = \alpha_0 + \alpha_1 \ln T_t + \alpha_2 \ln(WS/NI)_t + \alpha_3 \ln R + \alpha_4 Y_t + \varepsilon_t$ where C/M_2 is the ratio of currency holdings to broad money, Y is the real per capita income, R is the interest rate paid on time deposits, (WS/NI) is the ratio of wages and salaries in the national income, and T is an income tax variable.

series represents the excessive currency demanded as a result of the informal sector. The size of the informal economy, expressed as a percentage of the GDP, is then determined by multiplying the excessive currency by the velocity of money, which is assumed to be the same in both the formal and informal economy.

Table 1 provides basic summary statistics on the size of the informal sector, expressed as a percentage of the GDP, using the two data sources described above. According to the methodology based on Kaufmann and Kaliberda (1996) we estimate that the size of the informal sector during the 1990s is about 28 percent of the gross domestic product for the overall sample, 11 percent of the GDP for industrial countries, and almost 40 percent of the GDP in developing countries. Similarly, according to the estimates from Schneider and Klingmair (2003) the average size of the informal sector is 32 percent of the gross domestic product. The United States appears to be the country with the smallest informal sector, 9 percent of the GDP. The country with the largest informal sector is Bolivia, with 67 percent of the GDP. Among the industrial countries, the largest informal sector is in Greece with almost 30 percent of GDP. The average size of the informal sector in developing countries is 36 percent, and the smallest informal sector in this group of countries is in China and in Singapore, with about 13 percent of GDP. Both groups of estimates exhibit relatively large variations, and the simple correlation between the two data sets employed is 0.96 during the nineties.

INSERT TABLE 1 HERE

In particular, we use the following benchmark specification:

$$\text{Informal}_{\text{j}} = \alpha_0 + \alpha_1 \text{Ineq}_{\text{j}} + \alpha_2 \text{Inst}_{\text{j}} + \alpha_3 Rig_{\text{j}} + \alpha \ + X\beta + \varepsilon_j \qquad \text{for } j=1, 2,\ldots J \qquad (14)$$

where 'Informal' is the *dependent* variable and represents the size of the informal sector. Based on the theoretical model, our key explanatory variables are 'Ineq' which represents a measure of income distribution, and 'Inst' which is a measure of institutional quality. Additionally, we follow earlier literature and also include one previously explored variable, in particular, 'Rig' which is a measure of labor rigidities (e.g., Schneider and Klingmair, 2003; Johnson et al., 1998)[13]. Finally, *X* is a vector that includes basic macroeconomic controls, namely, output per capita, the rate of economic growth and the rate of inflation, an interactive dummy between inequality and institutions, and a developing country dummy. All the variables are for the beginning of the nineties.

With respect to the independent variables, we use the initial Gini coefficients as a proxy for income inequality from Deininger and Squire (1997). The Gini coefficient ranges from zero to one, and the higher the index the less equal the distribution of income[14]. According to Table 1, during the nineties the average Gini for the full sample is 0.39. In the case of industrial countries, the average Gini is 0.32, the country with the lowest income inequality in this group is Austria (0.23), and the one with the highest inequality is New Zealand (0.46). For developing and transition countries, the average Gini is about 0.40; the country with the lowest inequality is the Slovak Republic with 0.20, and the country with the highest one is Kenya with 0.61.

[13] We also included tax burden as an additional control (e.g., Cebula, 1997). This variable is typically positive and statistically significant and does not affect any of the other controls regardless of the method, proxy, or sample. We would be happy to provide these evidence upon request.

[14] While the data from Deininger and Squire (1997) go from 1960 to 1995 we are able to extend our inequality series using household data from Milanovic (2002a, 2002b) by generating information using the coefficient of variation of income and the income's linear correlation with ranks. For the sake of robustness, we also use alternative measures of income distribution such as the income share ratio of the top to the bottom quintile of the population as well as the income shares of the middle quintiles as well as other measures of inequality, in particular, Theil and Atkinson indices.

The institutional data come from International Country Risk Guide (ICRG) that has been often used in the literature, see, for example, Knack and Keefer (1995) and Hall and Jones (1999). The ICRG risk rating system assigns a numerical value to a predetermined range of risk components for a large number of countries We construct an average of the most commonly used institutional dimensions used in the literature, in particular, government stability, corruption, rule of law, democratic accountability, and bureaucracy quality. We also use these ICRG individual measures as well as the well-known Gastil (2002) data on civil liberties and political rights in order to test the robustness of our findings[15]. We rescaled these variables to 0-1, with higher scores implying more institutional freedom. Similarly, we also compute an aggregate index, defined as the simple average of the civil liberties and political rights indices.

Our source for labor market rigidities is the aggregate index of *de facto* labor regulations constructed in Forteza and Rama (2002). This index is the simple average of the ratio of the minimum wage to unit labor costs in the manufacturing sector, social security contributions as a percentage of salaries, total trade union membership as a percentage of total labor force, and the share of general government employment in total employment.[16] We use beginning-of-the-period data in the cross-section analyses.

As for the other controls employed in equation (14), we use the logarithm of the

[15] We also use data from Kaufmann et al. (2005) and find similar results. While the ICRG database is particularly useful in our cross-country regressions, its coverage is relatively limited when applying panel data regressions. In the latter case, we use Gastil's (2002) data which includes an annual assessment of the state of institutional freedom in each country and reports scores from 1 to 7, with lower scores denoting higher degrees of freedom.

[16] For robustness we follow Forteza and Rama (2002) and replicate the same exercise using a second index of regulations *de facto*, based on the simple average of the ratio of minimum wage to income per capita, the number of days maternity leave for a first child born without complications, the ratification of the ILO convention 87 that allows workers to establish organizations, and the ratio of central government employment to total employment. Results are very similar. Additionally, while limited for our purposes, we use data on labor cost by Heckman and Pages (2002) whenever possible. As before, the results do not

initial per-capita income; the average annual growth rate in gross domestic product per capita (Summers and Heston, 1991; World Bank, 2004) and the initial rate of inflation (International Monetary Fund, 2004).

5. RESULTS: CROSS-COUNTRY FINDINGS

The basic cross-country ordinary least squares results for the 1990s presented in Table 2 with all their obvious shortcomings, provide useful empirical regularities based on the predictions of the model[17]. We use our benchmark specification in (14) and use the two alternatives measures of the informal sector described in the previous section. Consistent with the theoretical model, the findings underscore the relevance of both income inequality and institutional quality as key determinants of informality.

INSERT TABLE 2 HERE

As predicted by Proposition 2, our findings in Table 2 indicate a positive and statistically significant link of income inequality and the size of the informal sector, regardless of the measurement method employed. For example, when income inequality measured by the Gini coefficient goes up from, say, Mexican levels (0.49) to Brazilian

change. We also use some limited data on entry costs whenever possible (La Porta et al., 2003). The variable yields a positive and statistically significant sign, which is not robust, though.

[17] Countries in the cross-country sample are Argentina, Australia, Austria, Belgium, Burkina Faso, Bangladesh, Bulgaria, Byelorussia, Bolivia, Brazil, Botswana, Canada, Switzerland, Chile, China, Cote d'Ivoire, Colombia, Costa Rica, Czech Republic, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Egypt, Spain, Finland, France, United Kingdom, Ghana, Greece, Guatemala, Hong Kong, Honduras, Croatia, Hungary, Indonesia, India, Ireland, Iran, Israel, Italy, Jamaica, Jordan, Japan, Kenya, Korea, Rep., Sri Lanka, Lithuania, Latvia, Morocco, Madagascar, Mexico, Mali, Malaysia, Niger, Nigeria, Nicaragua, Netherlands, Norway, New Zealand, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Senegal, Singapore, Slovak Rep., Sweden, Thailand, Tunisia, Turkey, Tanzania, Uganda,

levels (0.57) the relative size of the informal sector is predicted to increase by about 3 percentage points according to the demand for currency (Schneider and Klingmair, 2003) approach and by about 9 percentage points according to the Macroelectric approach (Kaufmann and Kaliberda, 1996). This is shown in regressions 1 and 3, respectively in Table 2. As implied by Proposition 1 above, the institutional variable is negatively and significantly correlated with the size of the informal sector regardless of whether it is estimated using the currency demand approach or the macroeconomic approach, as shown in column 1 and column 3. When an interactive term between the ICRG index and income inequality is introduced, the associated coefficient is negative and weakly statistically significant regardless of the dataset employed. This is shown in columns 2 and 4.[18] As for the other variables, the labor rigidities coefficient, while yielding the expected sign, is not statistically significant, which stands in contrast with most of the literature that typically finds labor rigidities to be an important determinant of informality.[19] These results do not change when using other labor cost measures (Heckman and Pages, 2001) as well as other cost of entrance variables (Djankov et al., 2001). Similarly, pairwise simple correlations between labor rigidity measures and proxies for the size of the informal sector, although positive, are close to zero. The other controls, such as the inflation rate in the nineties and a developing country dummy are not statistically significant in most specifications. However, the rate of economic growth

---

Ukraine, Uruguay, United States, Venezuela, RB., Vietnam, South Africa, Zambia, Zimbabwe.

[18] Note that when including the interactive term, the coefficient of the institutional measure turns out to be statistically insignificant implying that the interactive term between institutions and inequality is the more relevant one.

[19] We also included the coefficient of the tax rate which, as expected, is positive and statistically significant when using the currency demand approach; although it yields no statistical significance when using the Macroelectric approach. We do not include this variable in the benchmark regressions as taxes are also used to construct one of our proxies for the informal sector. However, in general, the tax burden is different in a developing country compared to an industrial country. In developing countries the indirect taxes and custom duties or other fees are more important than direct taxes. We are grateful to two

is consistently negative and statistically significant at five percent or higher. This, regardless of the informal sector proxy employed.

In addition to ordinary least squares we also try an instrumental variables approach. The instruments we consider are the average years of primary and secondary schooling attained by population with 25 or more years (Barro and Lee, 2001), the age dependency ratio defined as the number of people between 15 and 64 relative to the working population, the share of manufacturing in total output, religion, legal origin (World Bank, 2004), continental dummies, and a broad number of combinations of these variables as instruments[20]. Our results are shown in columns 5 to 8 in Table 2. In general, we obtain similar results to those using ordinary least squares. Specifically, the sign of the institutional variable remains negative and statistically significant as shown in regressions 5 and 7 in Table 2. The coefficient of the inequality variable is statistically significant and positive, although larger in most cases when compared to the ordinary least squares case. As before, the interactive term between institutions and inequality is always negative and statistically significant at ten percent or higher, as shown in regressions 6 and 8.

Our results are robust to changes in inequality measures, although more so when using the currency demand approach (Schneider and Klingmair, 2003), regardless of the econometric method, but less robust when considering the ordinary least squares version of the macroelectric approach (Kaufmann and Kaliberda, 1996). We show this in Table 3, where we employ a broad battery of inequality proxies, in particular, income quintiles

anonymous referees for this comment.

[20] Whereas we also tested beginning-of-the-period values for the explanatory variables and late values for the dependent variable, the endogeneity between inequality and informality are of obvious concern.

and income ratios[21]. For space considerations, only the coefficient of the inequality measure is reported which are based on our benchmark specification (14).

INSERT TABLE 3 HERE

Furthermore, our findings are quite robust to changes in institutional measures, as shown in Table 4. In fact, we obtain a negative and statistically significant coefficient at ten percent or higher, regardless of the institutional measure considered, such as aggregate governance (Kaufmann, et al, 2005), corruption, the rule of law, bureaucratic quality, government stability, democratic accountability (Knack and Keefer, 1995), and Gastil (1990). Interestingly, this finding holds regardless of the econometric method, and the inequality measure employed. In fact, the result holds when employing the benchmark specification (14) using the Gini coefficient as well as its instrumental variables version. Furthermore, the result holds when using income shares, or the ratio of income shares, instead. This is also shown in Table 4[22].

INSERT TABLE 4 HERE

6. FURTHER ROBUSTNESS ANALYSIS

A. CHANGES IN SPECIFICATION

We now apply Sala-i-Martín (1997) to test the robustness of our findings to changes in specification. To do this we look at the distribution of the estimator of the variable of

[21] We also test alternative inequality measures, such as Theil and Atkinson's indices. Overall we obtain similar results to those shown in Table 2. For the sake of economy we do not show these results but they are available upon request.

interest and focus on the fraction of the density function lying on each side of zero[23]. Given that zero divides the area under the density in two, we denote the larger of the two areas, *cdf(0)*, regardless of whether it is above or below zero. Under the assumption that the distribution of the coefficient of interest is non-normal the *cdf(0)* is calculated as follows. We consider a group of n variables classified as: (i) *dependent variable* as measured by our informality proxies, (ii) *core explanatory variables* based on our benchmark specification in (14), and (iii) *ancillary variables* representing a set of related auxiliary variables identified as related to determination of the size of the informal sector. Using our benchmark specification we augment our empirical models by using the pool of ancillary variables. The idea is to choose up to three variables at a time, and perform regressions using all the possible combinations based on our pool of ancillary variables[24]. We test our basic specification for all our possible combinations of ancillary variables and compute the coefficient estimates, its variance, the integrated likelihood, and the individual *cdf(0)* for each regression. We compute the aggregate *cdf(0)* of our coefficient of interest as the weighted average of all individual *cdf(0)s* where the weights are the integrated likelihoods as defined in Sala-i-Martin 1997) The variable of interest is said to be strongly correlated (i.e., is robust) if the weighted *cdf(0)*, is greater than or equal to 0.95.

INSERT TABLE 5 HERE

---

[22] When using Theil and Atkinson inequality measures we also obtain robust results.

[23] If 95 percent of the density function for the estimates of the coefficient of interest lies to the right of zero, one could say that this variable is more likely to be correlated with our dependent variable.

[24] Our set of ancillary variables includes fourteen variables, most of them from the World Bank (2004) except the institutional variables which are from ICRG (2005). In particular, we considered the unemployment rate, the labor participation rate, four regional dummies, regulatory burden, business and entry regulation, tax rate, gross domestic product in the official economy, percentage working part-time, primary education, and secondary education. We are grateful to anonymous referees for suggesting several

Our findings are shown in Table 5. That is, we test for robustness for each of the variables in the regression one at a time using our benchmark regression in (14). The informality proxy employed is the one from Schneider and Klingmair (2003)[25]. In the first row of each corresponding variable in Table 5 we report the aggregate *cdf(0)* under the assumption of non-normality. The second row presents the standard deviation computed as the squared root of the weighted variance estimate for all the regressions. In general, we find that most of the variables included in the regressions are robust at five percent or higher. For instance, this is the case of our cross-country specification when our inequality variable is the Gini coefficient, regardless of the institutional variable employed. This is shown in the first row on Table 5. Similarly, we find that our findings are also robust to changes in specification when applying instrumental variables, and regardless of the inequality measure employed, in particular, income shares and income ratios.

B. PANEL DATA ANALYSIS

We next assemble a panel data set of industrial and developing countries, which consists of at most six non-overlapping 5-year period observations over the sample period 1970 to 2000[26]. The premise here is that changes in inequality occur relatively slowly so that the

---

of the controls above.

[25] The results using the macroelectric approach are very similar and consequently are not reported. We would be happy to provide them.

[26] Countries in the sample are United Arab Emirates, Argentina, Australia, Austria, Belgium, Burkina Faso, Bangladesh, Bulgaria, Bahrain, Bahamas. Byelorussia, Bolivia, Brazil, Botswana, Canada, Switzerland, Chile, China, Cote d'Ivoire, Colombia, Costa Rica, Cyprus, Czech Republic, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Egypt, Arab Rep., Spain, Estonia, Ethiopia, Finland, France, United Kingdom, Ghana, Guinea, Guinea Bissau, Greece, Guatemala, Hong Kong, Honduras, Croatia, Hungary, Indonesia, India, Ireland, Iran, Iraq, Israel, Italy, Jamaica, Jordan, Japan, Kazakhstan, Kenya, Kyrgyz Republic, Korea, Rep., Kuwait, Lebanon, Libya, Sri Lanka. Lesotho, Lithuania, Luxembourg,

observed time variation is rather small. Because of considerable data limitations, in this section we are only able to use the Macroelectric approach. We employ GMM-IV techniques (Arellano and Bover, 1995; Arellano-Bond, 1991) to minimize endogeneity problems. By using this method we estimate a regression equation in differences and a regression equation in levels simultaneously, with each equation using its own specific set of instrumental variables. The consistency of the GMM estimator depends on whether lagged values of the explanatory variables are valid instruments in the regression. We address this issue by considering two specification tests[27].

For the sake of comparison, we first use fixed effects ordinary least squares regressions, which are presented in Table 6. Overall, the results are very similar to those from the cross-section regressions[28]. In particular, the coefficient of the inequality variable is positive and statistically significant at five percent, while the interactive term between institutions and inequality is negative, although it is now statistically significant

---

Latvia. Morocco, Madagascar, Mexico, Mali, Malta, Mongolia, Mauritania, Mauritius, Malaysia, Niger, Nigeria, Nicaragua, Netherlands, Norway, Nepal, New Zealand, Oman, Pakistan, Panama, Peru, Philippines, Papua New Guinea, Poland, Portugal, Paraguay, Qatar, Romania, Russian Federation, Rwanda, Saudi Arabia, Senegal, Singapore, Sierra Leone, El Salvador, Slovak Republic, Slovenia, Sweden, Syria, Thailand, Trinidad and Tobago, Tunisia,, Turkey, Taiwan, Tanzania, Uganda, Ukraine, Uruguay, United States, Venezuela, Vietnam, Yemen, Yugoslavia, South Africa, Zambia, Zimbabwe. In order to maximize the time-span of the panel we use the Gastil data. ICRG has too few usable observations.

[27] The first is a Sargan test of over-identifying restrictions, which tests the overall validity of the instruments. Failure to reject the null hypothesis gives support to the model. The second test examines the hypothesis that the error term is not serially correlated. We test whether the differenced error term is second, and third order serially correlated. Notice that first-order serial correlation is expected by construction. Second-order serial correlation of the differenced residual indicates that the original error term is serially correlated and follows a moving average process at least of order one. If the test fails to reject the null hypothesis of absence of second-order serial correlation, the original error term is serially uncorrelated.

[28] One may have serious concerns about the use of panel data with indicators of informality based on the proxies employed in this paper. Consider electricity, for example. In countries where many power plants are built, electricity demand tends to increase because production technologies become more electricity intensive and not necessarily because informal activity increased. Similarly, when a country develops, the financial system develops as well, and the economy as an aggregate may become less barter based. These effects may be even more relevant in developing countries, where technologies and financial systems evolve quickly. This may be the reason why evidence with cross-country regressions may be more sensible and panel approaches are less used even if they are technically possible to apply. Given the fact that cross-country variation is larger than time variation, and that heuristically informality indicators are sensible

at ten percent only[29]. Notice that unlike in the cross-section case, the coefficients of labor rigidities yields the expected sign at statistically significant levels.

INSERT TABLE 6 HERE

We also present within group estimators that, likewise, ignore endogeneity and may be biased in short panels. This is shown in columns 3 and 4 in Table 6. Overall, we obtain virtually the same results as in the pooled fixed effects case, that is, a positive coefficient in the inequality term, and a negative coefficient in the interactive term. We also obtain a positive and statistically significant sign in the labor rigidities variable and in the interactive term between inequality and institutions (Regression 4). Furthermore, we present level by level estimators, instrumented by lagged values which also have somewhat similar problems as the previous case. Still, when applying this method, our estimates are quite consistent with our previous findings. This is shown in columns 5 and 6 in the same table. Finally, when applying a dynamic panel data with GMM-IV, we find that, as in the other cases, the coefficient of the Gini index is positive and statistically significant in the two specifications considered. We also find that the sign of the interactive term is negative, although only weakly significant in Regression 2, whereas the institutional variable is not statistically significant. As to the other controls, we obtain the expected and statistically significant signs for output per capita, rate of economic growth, and the inflation rate, as well as with respect to labor rigidities.

---

across countries, simple cross-country regressions seem more adequate. Our thanks to an anonymous referee for this comment.

[29] Also, notice that the coefficient of the institutional variable is not statistically significant when excluding the interactive term.

7. CONCLUSIONS

This paper presents theory and evidence on the relationship between inequality and informality. We propose a simple theoretical model where the increase in income inequality, by lowering the relative benefits from becoming formal for the poor, causes a bigger informal sector, more so the weaker the institutions. We test our theory using different proxies for the size of the informal sector, income inequality, and institutional quality and employing a broad range of econometric techniques in a cross-country sample and a panel analysis. Overall, the empirical findings are consistent with the basic predictions of the model, in particular regarding (i) the positive link between income inequality and the size of the informal sector, and (ii) the interaction between institutions and inequality in their impact on informality. Additionally, the evidence shows that some of the commonly believed determinants of informality are not robust in all specifications and lose their significance in the presence of the above factors.

Table 1. Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A					
Currency Demand --Schneider and Klingmair	99	0.315	0.136	0.087	0.671
Macroelectric --Kaufmann and Kaliberda	68	0.254	0.169	-0.082	0.686
Gross Domestic Product per capita	110	8.533	1.072	6.174	10.322
Rate economic growth	110	0.015	0.025	-0.076	0.075
Inflation rate	115	0.115	0.116	0.003	0.575
Gini coefficient	107	0.383	0.092	0.209	0.585
ICRG Average Index	116	4.334	0.910	2.043	6.398
Labor rigidity	114	0.296	0.174	0.017	0.802
Income share Bottom 20	96	0.065	0.022	0.019	0.118
Income share Bottom 40	96	0.172	0.048	0.074	0.274
Income share Middle 20	96	0.158	0.029	0.092	0.251
Income share Top 20	96	0.458	0.088	0.316	0.653
Income share Top 40	96	0.669	0.071	0.544	0.829
Ratio of income shares Top 20 / Bottom 20	96	9.141	6.754	2.702	35.191
Ratio of income shares Top 20 / Bottom 40	96	4.448	2.162	1.986	11.014
Atkinson's Inequality (n=1)	103	0.789	0.022	0.753	0.842
Theil coefficient	103	-0.267	0.029	-0.316	-0.194
Least developed country dummy	122	0.820	0.386	0.000	1.000
Age dependency ratio	122	0.725	0.185	0.459	1.092
% of modern sector	114	0.846	0.135	0.437	0.998
% of Catholic to total pop. (1980)	121	32.353	37.332	0.000	97.300
% of Muslim to total pop. (1980)	121	24.483	36.800	0.000	99.500
British legal origin	121	0.289	0.455	0.000	1.000
French legal origin	121	0.463	0.501	0.000	1.000
Socialist legal origin	121	0.165	0.373	0.000	1.000
German legal origin	121	0.050	0.218	0.000	1.000
Scandinavian legal origin	121	0.033	0.180	0.000	1.000
Latitude	121	0.321	0.187	0.011	0.711
Panel B					
(i) Industrial Countries					
Gini coefficient		0.316	0.051	0.227	0.456
Currency Demand --Schneider and Klingmair		0.169	0.062	0.087	0.286
Marcoelectric --Kaufmann and Kaliberda		0.111	0.131	0.075	0.362
(ii) Developing Countries					
Gini coefficient		0.404	0.091	0.203	0.606
Currency Demand --Schneider and Klingmair		0.355	0.111	0.131	0.671
Macroelectric --Kaufmann and Kaliberda		0.396	0.200	0.192	0.691

TABLE 2. CROSS-SECTION: ORDINARY LEAST SQUARES AND INSTRUMENTAL VARIABLES, 1990-2000

	Ordinary Least Squares				Instrumental Variables			
	Currency Demand Schneider and Klingmair (2003)		Macroelectric approach Kaufmann and Kaliberda (1996)		Currency Demand Schneider and Klingmair (2003)		Macroelectric approach Kaufmann and Kaliberda (1996)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP per capita	-0.007	-0.006	-0.006	-0.005	0.014	0.014	0.021	0.013
	(0.023)	(0.023)	(0.023)	(0.024)	(0.034)	(0.033)	(0.034)	(0.034)
Economic growth	-1.257	-1.113	-1.369	-1.269	-1.674	-1.328	-1.611	-1.300
	(0.431)***	(0.447)**	(0.450)***	(0.481)**	(1.191)	(1.227)	(1.058)	(1.123)
Inflation rate	0.134	0.164	0.091	0.101	-0.464	-0.407	-0.481	-0.388
	(0.096)	(0.095)*	(0.106)	(0.107)	(0.233)*	(0.227)*	(0.173)***	(0.154)**
Gini coefficient	0.267	1.324	0.337	1.164	0.570	2.595	0.270	2.706
	(0.124)**	(0.605)**	(0.153)**	(0.844)	(0.247)**	(1.140)**	(0.169)*	(0.845)***
ICRG Index	-0.091	0.001	-0.088	-0.017	-0.123	0.052	-0.147	0.062
	(0.027)***	(0.050)	(0.026)***	(0.062)	(0.035)***	(0.100)	(0.034)***	(0.073)
Labor rigidities	0.098	0.075	0.092	0.085	0.044	-0.007	-0.000	-0.068
	(0.072)	(0.069)	(0.076)	(0.076)	(0.165)	(0.153)	(0.153)	(0.143)
Gini*ICRG Index		-0.255		-0.200		-0.511		-0.573
		(0.137)*		(0.115)*		(0.282)*		(0.192)***
LDC dummy	-0.006	0.003	-0.006	0.005	0.030	0.044	0.029	0.030
	(0.027)	(0.027)	(0.030)	(0.029)	(0.058)	(0.052)	(0.052)	(0.047)
Constant	0.659	0.252	0.618	0.298	0.542	-0.167	0.711	-0.126
	(0.166)***	(0.272)	(0.187)***	(0.380)	(0.260)**	(0.404)	(0.231)***	(0.302)
Observations	86	86	83	83	62	62	65	65
R-squared	0.68	0.69	0.67	0.67	0.56	0.61	0.59	0.63

(*) statistically significant at ten percent; (**) statistically significant at five percent. Standard errors are in parentheses.

TABLE 3. CROSS-SECTION: CHANGES IN INEQUALITY MEASURES, 1990-2000

	Ordinary Least Squares		Instrumental Variables	
	Currency Demand Schneider and Klingmair (2003)	Macroelectric approach Kaufmann and Kaliberda (1996)	Currency Demand Schneider and Klingmair (2003)	Macroelectric approach Kaufmann and Kaliberda (1996)
	(1)	(2)	(3)	(4)
I. Income shares				
Top 20	0.374	0.426	0.362	0.746
	(0.137)***	(0.202)**	(0.162)**	(0.239)***
Top 40	0.418	0.318	0.446	0.626
	(0.167)**	(0.257)	(0.201)**	(0.287)**
Middle 20	-0.589	-0.410	-1.235	-1.135
	(0.417)	(0.622)	(0.614)**	(0.795)
Bottom 40	-0.682	-0.533	-0.670	-1.126
	(0.232)***	(0.339)	(0.300)**	(0.448)**
Bottom 20	-1.324	-0.805	-1.442	-1.726
	(0.455)***	(0.644)	(0.619)**	(0.874)*
II. Ratio of income shares				
Top 20 / Bottom 20	0.005	0.005	0.005	0.009
	(0.002)***	(0.002)**	(0.002)**	(0.003)***
Top 20 / Bottom 40	0.015	0.015	0.016	0.026
	(0.005)***	(0.008)*	(0.007)**	(0.009)***

(*) statistically significant at ten percent; (**) statistically significant at five percent. Standard errors are in parentheses.

TABLE 4 CROSS-SECTION: CHANGES IN INSTITUTIONAL MEASURES, 1990-2000

	Governance	Corruption	Rule of Law	Bureaucratic Quality	Government Stability	Democratic Accountability	Gastil Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cross-Section	-0.104 **	-0.055 **	-0.059 **	-0.049 **	-0.051 **	-0.044 **	0.454**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.18)
Instrumental Variables	-0.107 **	-0.050 **	-0.062 **	-0.044 **	-0.063 **	-0.036 *	0.455**
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.18)
I. Income Shares							
Top20	0.407 **	0.540 **	0.475 **	0.530 **	0.534 **	0.510 **	0.565 **
	(0.14)	(0.15)	(0.16)	(0.16)	(0.15)	(0.16)	(0.22)
Top40	0.520 **	0.677 **	0.565 **	0.637 **	0.601 **	0.675 **	0.729 **
	(0.17)	(0.19)	(0.20)	(0.19)	(0.20)	(0.18)	(0.30)
Middle 20	-0.880 **	-1.149 **	-0.853 *	-1.106 **	-0.867 *	-1.247 **	-2.168 **
	(0.42)	(0.47)	(0.50)	(0.49)	(0.54)	(0.44)	(0.90)
Bottom 40	-0.843 **	-1.098 **	-0.955 **	-1.032 **	-1.029 **	-1.061 **	-1.085 **
	(0.26)	(0.29)	(0.29)	(0.31)	(0.29)	(0.28)	(0.44)
Bottom 20	-1.525 **	-2.019 **	-1.644 **	-1.822 **	-1.721 **	-1.998 **	-2.383 **
	(0.45)	(0.53)	(0.53)	(0.55)	(0.57)	(0.50)	(0.98)
II. Ratio of Income Shares							
Top20 / Bottom 20	0.005 **	0.007 **	0.006 **	0.007 **	0.006 **	0.007 **	0.009 **
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Top20 / Bottom 40	0.017 **	0.023 **	0.019 **	0.022 **	0.021 **	0.022 **	0.025 **
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

(*) statistically significant at ten percent; (**) statistically significant at five percent. Standard errors are in parentheses.

TABLE 5. ROBUSTNESS TO CHANGES IN SPECIFICATION, 1990-2000

	ICRG Index (1)	Corruption (2)	Rule of Law (3)	Bureaucratic Quality (4)	Government Stability (5)	Democratic Accountability (6)	Gastil Index (7)
Cross-Section	0.99	0.99	0.99	0.97	0.98	0.99	0.95
	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)
Instrumental Variables	0.99	0.99	0.99	0.97	0.97	0.98	0.94
	(0.04)	(0.02)	(0.03	(0.02)	(0.03)	(0.03)	(0.05)
. *Income Shares*							
Top20	0.98	0.98	0.98	0.97	0.96	0.99	0.95
	(0.11)	(0.12)	(0.13	(0.17	(0.15)	(0.13	(0.11)
Middle 20	0.98	0.97	0.99	0.97	0.97	0.98	0.96
	(0.51)	(0.51)	(0.48)	(0.49)	(0.53)	(0.49)	(0.37)
Bottom 20	0.97	0.99	0.99	0.98	0.95	0.95	0.95
	(0.35)	(0.45)	(0.52)	(0.47)	(0.62)	(0.60)	(0.58)
II. Ratio of Income Shares							
Top20 / Bottom 20	0.97	0.98	0.98	0.98	0.96	0.99	0.94
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Top20 / Bottom 40	0.98	0.98	0.98	0.97	0.96	0.98	0.92
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)

The second row presents the standard deviation of the variable of interest while the first row shows the cumulative distribution function (0). Our variable of interest is the size of the informal sector using the currency demand method (Schneider and Klingmair, 2003). A variable whose weighted *cdf(0)* is larger than 0.95 is significantly correlated with the dependent variable (i.e. robust) at a 5 % significance level. The *cdf* is computed assuming non-normality of the parameters estimated. Results are similar if we assume normality, instead. The benchmark regression employed is that in specification (14) on the text.

TABLE 6: PANEL DATA ANALYSIS, 1970-2000

	Panel OLS with Fixed Effects		Within Groups Fixed Effects		Dynamic Panel Data		Dynamic Panel Data GMM-IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.349 **	0.135	0.158	-0.171	0.653 **	0.875 **	0.779 **	0.896 **
	-0.16	-0.19	-0.16	-0.25	-0.08	-0.14	-0.09	-0.15
Output per capita	-0.029	-0.038 *	-0.028	-0.031	-0.061 **	-0.07 **	-0.059 **	-0.08 **
	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02
Economic Growth	-0.188	-0.134	-0.217	0.198	-0.136 **	-0.161 *	-0.123 **	-0.182 *
	-0.44	-0.43	-0.45	-0.44	-0.06	-0.1	-0.06	-0.11
Inflation Rate	-0.009	0.034	-0.013	0.004	0.099 **	0.066 *	0.094 **	0.083 *
	-0.09	-0.08	-0.09	-0.09	-0.05	0.04	-0.04	-0.05
Gini Coefficient	0.586 **	1.313 **	1.000 ***	1.973 **	0.194 **	0.261 **	0.172 **	0.236 **
	-0.13	-0.42	-0.19	-0.56	-0.09	-0.1	-0.09	-0.1
Gini * Institutions	…	-1.146 *	…	-1.483 *		-1.322 *		-1.265 *
		-0.61		-0.77		-0.71		-0.75
Institutions	0.042	0.497 *	0.067	0.61 *	-0.139	-0.047	-0.109	-0.048
	-0.06	-0.26	-0.06	-0.32	-0.16	-0.14	-0.13	-0.14
Labor Rigidities	0.243 **	0.217 **	0.194 **	0.175 *	0.324 **	0.316 **	0.337 **	0.292 **
	-0.1	-0.1	-0.1	-0.1	-0.06	-0.09	-0.06	-0.08
Observations	417	417	381	381	283	283	283	283
R-Squared	0.1426	0.1541	0.1671	0.178	0.373	0.371	0.373	0.371
Sargan Test					0.715	0.695	0.893	0.872
2nd Order Correlation					0.267	0.269	0.301	0.326
3rd Order Correlation					0.326	0.315	0.318	0.267

(*) statistically significant at ten percent; (**) statistically significant at five percent; (***) statistically significant at one percent. Standard errors are in parentheses. Sargan and correlation tests are p-values. Dynamic panel data refers to the method by Arellano-Bond (1991). Dynamic Panel Data GMM-IV refers to the method by Arellano-Bover (1995).

REFERENCES

Acemoglu, D. (1995) "Reward Structure and the Allocation of Talent," *European Economic Review*, 39, 17-33.

Acemoglu, D. and T. Verdier (1998) "Property Rights, Corruption, and the Allocation of Talent: A General Equilibrium Approach," *Economic Journal*, 108: 1381-1403.

Arellano, Manuel and Stephen Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations" *Review of Economic Studies*, *58:* 277-297

Arellano, M. and O. Bover (1995) "Another look at the Instrumental Variable Estimation of Error-Component Models," *Journal of Econometrics* 68, 29-51.

Barro, R. and J. Lee (1993) "International Comparisons of Educational Attainment", *Journal of Monetary Economics*, 32, 363-394.

Botero, J., S. Djankov, R. La Porta, F. Lopez-de-Silanes, A. Shleifer (2004) "The Regulation of Labor," forthcoming, *Quarterly Journal of Economics.*

Cebula, R.J., 1997, "An empirical analysis of the impact of government tax and auditing policies on the size of the underground economy: The case of the United States, 1973-94," *The American Journal of Economics and Sociology*, 56, 73-186.

Chong, A. and M. Gradstein (2004) "Inequality and Institutions," Manuscript, Inter-American Development Bank.

Djankov, S., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2001) "The Regulation of Entry," NBER Working Paper.

Deininger, K. and L. Squire (1997) "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10, 3: 565-91.

Forteza, A. and Rama, M. (2001) "Labor Market Rigidity and the Success of Economic Reforms Across More than 100 Countries," Working Paper, The World Bank.

Gastil, R. (1990) "The Comparative Survey of Freedom: Experiences and Suggestions". *Studies in Comparative International Development*, *25, 1:* 25-50.

Giles, D., E.A. (1999) "Modeling the hidden economy in the tax-gap in New Zealand," *Working paper*, Department of Economics, University of Victoria.

Giles, D., E.A. and Tedds, L.M. (2002) "Taxes and the Canadian underground economy," Canadian Tax Paper 106, Canadian Tax Foundation, Toronto.

Hall, R. and C. Jones (1999), "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, 114, 83-116.

Heckman, J. and C. Pages (2001) "Regulation and Deregulation: Lessons from Latin American Labor Markets", *Economia*, 1, 123-145.

International Country Risk Guide (2005). http://www.icrg.org

International Monetary Fund (2004) International Financial Statistics, http://www.imf.org

Kaufmann, D., A. Kraay, M. Mastruzzi, (2005). "Governance Matters III: Governance Indicators for 1996-2002," World Bank Policy Research Working Paper 3106.

Kaufmann, D. and A. Kaliberda, (1996): "Integrating the Unofficial Economy into the Dynamics of the Post Socialist Economies: A Framework of Analyses and Evidence," World Bank Policy Research Working Paper 1691.

Knack, S. and P. Keefer (1995), "Institutions and economic performance: Cross-country tests using alternative institutional measures," *Economics and Politics* 7: 207-227.

KPMG (2003) Data on Corporate Tax around the World: http://www.kpmg.com

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1998) "Law and Finance", *Journal of Political Economy*, 106, 1113-1155.

Lacko, M. (1996), "Hidden Economy - An Unknown Quantity? Comparative Analysis of Hidden Economies in Transition Countries in 1989-1995," Working Paper 9905, Economics Department, University of Linz, Austria.

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., A Oviedo, and L. Serven (2005) "The Impact of Regulation on Growth and Informality: Cross-Country Evidence", AEI-Brookings Joint Center for Regulatory Studies, Related Publications 05-11, Washington, DC.

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

Milanovic, B. (2002a), "True World Income Distribution, 1988 and 1993: First Calculation based on Household Surveys Alone," *Economic Journal*, 112, 51-92.

Milanovic, B. (2002b), "Worlds Apart: Inter-National and World Inequality 1950-2000," Washington, DC: The World Bank.

Rama, M. (1995) "Do Labor Market Policies and Institutions Matter? The Adjustment Experience in Latin America and the Caribbean," *Labour*, S243-S268.

Saavedra, J. and A.Chong (1999) "Structural Reforms, Institutions and Earnings: Evidence From the Formal and Informal Sectors in Urban Peru", *Journal of Development Studies*, 35, 95-116.

Schneider, F. and D. Enste, 2000, "Shadow Economies: Size, Causes and Consequences," *Journal of Economic Literature*, 38, 77-114.

Schneider, F. and R. Klingmair (2003) "Shadow Economies around the World: What do We Know?" Manuscript, University of Linz, Austria.

Schneider F. (2005) "Shadow Economies of 145 Countries All Over the World: What Do We Really Know? Manuscript Johannes Kepler University of Linz.

Summers, R. and A. Heston (1991) "The Penn World Table (Mark V): An Expanded Set of International Comparisons, 1950-1988, *Quarterly Journal of Economics*, 106, 2: 327-68.

Tanzi, V., 1999, "Uses and abuses of estimates of the underground economy," *Economic Journal*, 109, F338-47.

Ziliak, J. (1997) "Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business Economics and Statistics*, 15 (4).

World Bank (2004) World Development Indicators, http://www.worldbank.org