Informality, Firm Size and Economic Growth: Testing the de Soto Hypothesis*

Lennart Erickson**

Brown University

Preliminary Draft: October 8, 2002

ABSTRACT: Formalizing a concept developed by de Soto (1989), I construct a simple general equilibrium model of the informal sector. Agents in the economy can choose to be entrepreneurs in the formal or informal sector, or work in the labor market. Government rent-seeking and bureaucracy leads to a bifurcation in the size of firms: small informal firms and large formal firms. The empirical implications of the model are tested using cross-country data. The data do not conform to the model's predictions, indicating that this popular conception of the nature of informal sectors should be re-examined.

Keywords: Informal sector, firm size.

*I am grateful to David Weil for his comments and support. Thanks also to James Feyrer, Michal Jerzmanowski, Deitrich Vollrath and the participants of the Brown Macrolunch.

**Dept. of Economics, Box B, Brown University, Providence, RI 02912, Lennart_Erickson@brown.edu

1. Introduction

In 1989, the English edition of *The Other Path: the Invisible Revolution in the Third World*, by the Peruvian social scientist Hernando de Soto, was published to great acclaim. De Soto argued that a key factor contributing to poverty in the developing world were the barriers placed by governments in the path of small-scale entrepreneurs. His story was straightforward: bureaucrats, either through needless licensing requirements or outright corruption, make setting up a legal business very costly. Not surprisingly, many would-be entrepreneurs are dissuaded from starting legal businesses in such conditions. Instead, they set up "informal" firms. These firms do not have the required permits and operate largely outside the reach of the government. However, they do so at a cost of productivity, lacking access to infrastructure, credit markets and legal institutions. It is important to note that the activity these firms engage in is not intrinsically criminal, *e.g.,* drug dealing, but rather they are producing legal goods and services outside of regulatory norms. Other entrepreneurs operate in the "formal" sector. While not necessarily complying all legal rules, these firms are not hiding from the government. They incur costs in the form of taxes and bureaucratic compliance.

To illustrate how costly these obstacles could be, de Soto had a team of researchers try to legally register a fictitious small garment factory in Lima, Peru. In order to comply with all the necessary legal requirements, the team had to obtain eleven different permits. From start to finish, the process took 289 days and cost, in terms of fees and forgone profits, thirty-two times the monthly minimum

wage. (de Soto, p. 133)

A key part of this story is firm size. Large firms operate in the formal sector, where they can reap the combined fruits of efficiency from both legality and scale. Very small firms operate in the informal sector. They sacrifice efficiency from scale and legality, but avoid government rent-seeking.

> When we visited clandestine factories, we found that the need to avoid detection forces informals to operate on a very small scale. The deliberately limit their operations or, if they need to grow, do so by dispersing their workers so that there are never more than ten in one establishment. While such arrangements obviously help them avoid detection, they also prevent them from achieving efficient scales of production. This seems to be a fairly widespread consequence of informality. (de Soto, p. 153)

Higher levels of government rent seeking or bureaucratic obstacles to legal firm entry will lead to a greater bifurcation of firm sizes; very small informal firms, large formal firms and an absence of small and medium sized formal firms. It is this distortion in firm size distribution on which this paper focuses. Using data from on manufacturing establishment size distributions from the United Nations Industrial Development Organization (UNIDO) and data on bureaucratic barriers to entry from Djankov, et al. (2000), I examine whether this bifurcation is associated with government rent seeking. Since cross-country data on informal firm size distributions are unavailable, I consider the general equilibrium effect which the

existence of informal sectors should have on the number and size of formal sector establishments. Using a number of different specifications, I fail to find evidence of this distortion borne out empirically.

The broader goal of this paper is to contribute to the economic growth literature on cross-country productivity differences. There is a growing consensus in that literature that while Solow-style explanations can account for much of the variation in growth between countries, (Mankiw, et al. (1992), there remain unexplained differences in total factor productivity. (Hall and Jones (1999), Klenow and Rodriguez-Claire (1997)) Therefore, exploring the nature of these productivity differences should be paramount. Mauro (1995), an example of this line of research, finds that high corruption is correlated with low growth. The de Soto hypothesis offers a very intuitive channel for how corruption, through the distortions of informality, could lead to lower TFP. If, as appears to be the case, informality does not follow the pattern implicit in the de Soto hypothesis, then the story's implications for growth differences must be reconsidered.

The paper proceeds as follows. Section two reviews the related literature. Section three presents a simple general equilibrium model and considers simulation results deriving from that model. Section four tests those implications empirically. Section five concludes.

2. Literature Review

The question of informal production has received sustained interest over the last thirty years, starting with implicitly with Harris and Todaro(1970). Much effort has been made to estimate the size of the "unofficial" or "shadow" economy. Usually, these formulations include what de Soto and this paper would describe as the informal sector, as well as explicitly illegal activity and unreported economic activity by legally registered firms. Three main approaches have been used. First, Schneider (1997) and Williams and Windebank (1995) use the prevalence of cash in monetary transactions as a proxy for unofficial activity. Lacko (1996) estimates unoffical activity through the over consumption of household electricity. Finally, Loayza (1997) estimates unofficial production using a variety of explanatory variables including labor market restrictions, corruption, political risk and rule of law. Freidman, et al. (2000), using data on the size of the unofficial economy from these and other sources, find that while higher tax rates are associated with less unofficial activity, corruption is associated with more unofficial activity.[1]

A variety of authors have proposed models of informal and unofficial activity, including Azuma and Grossman (2002), Choi and Thum (2002), Sarte (2002) Bello (1999), Fukuchi (1998) and Marcouiller and Young (1995). However, none of these are explicitly concerned with firm size distribution. Yamada (1996) develops a model of the informal sector and compares it to household survey data in Peru. Fortin, et al., (1997), builds a computable general equilibrium model of

[1] For a comprehensive overview of the unofficial economy literature, see Scheinder and Enste (2000).

the informal sector in Cameroon.

In addition to de Soto, my model draws primary inspiration from two other works. Lucas (1978) presents a general equilibrium model of firm size. In an economy with fixed labor and capital, individual workers are endowed with varying levels of entrepreneurial skill. Individuals with high skill form firms and hire less skilled individuals as workers. Entrepreneurs with greater skill have larger firms. Rauch (1992) modifies Lucas' approach to model the informal sector. He eliminates capital and imposes a fixed, above-market-clearing wage for firms operating in the formal sector. Less skilled entrepreneurs operate in the informal sector. I depart from Rauch's approach assuming a single wage across the two sectors (more on this assumption below). In my model, entrepreneurs face a trade off of higher productivity and rent payment in the formal sector versus lower productivity and rent avoidance in the informal sector. This approach is more faithful to de Soto's original story and also allows for better consideration of productivity effects arising from informal production.

3. The Model

There is a continuum of individuals in the economy. Each individual is endowed with a level of entrepreneurial skill, x, which is distributed uniformly over the interval $[0, 1]$ There is no capital. Individuals can choose to

1) Start a firm in the formal sector.

2) Start a firm in the informal sector.

3) Work for a wage, w, which is the same in either sector.

Individuals with the most entrepreneurial skill will start firms in the formal sector, while those with less entrepreneurial skill will start informal firms. Individuals with the lowest levels of entrepreneurial skills will enter the labor market. An equilibrium wage clears the labor market.

3.1. Single Labor Market

The model assumes that the wage, w, is the same in both the informal and formal sector. This is a major assumption, as much of the theoretical literature and empirical studies describe an above-market-clearing wage in the formal sector; indeed, this is the approach adopted by Rauch. However, the use a single wage does have some important advantages.

A single wage makes the model much more tractable. With excess demand for formal sector-labor, individuals would have to evaluate the probability of working in the formal or informal sector and some account would have to be given of the sorting process. This would complicate the model without adding significant predictive properties. Also, the idea of a single fixed wage seems implausible over time and across countries. A more plausible formulation would be some premium over the market-clearing wage, but this approach too would not add much to the empirical predictions of the model.

Moreover, recent empirical work has cast doubt on the dualistic labor market formulation. Maloney (1998) uses wage data from Mexico to challenge the existence of a formal sector wage premium all together.

3.2. Formal Production

An entrepreneur in the formal sector can earn a profit of

$$xn_F^{\alpha} - wn_F - \tau \tag{1}$$

where x is the individual's entrepreneurial skill, α is a production coefficient such that $\alpha \subset [0,1]$, w is the wage at which the entrepreneur can hire labor, τ is a fixed rent or tax which must be paid to the government. The choice variable for the entrepreneur is n_f, the number of workers she hires. The entrepreneur hires the amount of workers which maximizes her profits. The first order condition is,

$$\alpha x n_F^{\alpha-1} - w = 0 \tag{2}$$

which yields the optimal amount of workers for the firm of

$$n_F^* = x^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w}\right)^{\frac{1}{1-\alpha}} \tag{3}$$

3.3. Informal Production

An entrepreneur in the informal sector can earn of

$$(1-\theta)\, xn_I^{\alpha} - wn_I \tag{4}$$

where n_I is the amount of workers hired and $\theta \subset [0,1]$ is a parameter which measures the efficiency loss of being in the informal sector. The first order

condition for profit maximization is

$$(1-\theta)\,\alpha x n_I^{\alpha-1} - w = 0 \tag{5}$$

yielding the optimal amount of workers to hire of

$$n_I^* = x^{\frac{1}{1-\alpha}} \left(\frac{(1-\theta)\alpha}{w} \right)^{\frac{1}{1-\alpha}} \tag{6}$$

3.4. Formal/Informal Boundary

An individual will choose whether to be an entrepreneur in the formal or informal sector depend on the profits she can make in each. For the individual who is indifferent between the two options, who I will denote at the individual with $\overline{x}$ level of entrepreneurial skill, the profit she can achieve in either sector will be the same.

$$\overline{x} n_F^{*^\alpha} - w n_F^* - \tau = (1-\theta)\,\overline{x} n_I^{*^\alpha} - w n_I^* \tag{7}$$

Substituting for n_F^* and n_I^* yields

$$\overline{x} = \tau^{1-\alpha} w^\alpha \left[\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}} \right) \left(1 - (1-\theta)^{\frac{1}{1-\alpha}} \right) \right]^{\alpha-1} \tag{8}$$

3.5. Informal/Working Boundary

Similarly, there will be an individual $\underline{x}$ who will be indifferent between being an informal sector entrepreneur and working.

$$(1-\theta)\,\underline{x}n_I^{*^{\alpha}} - wn_I^* = w \tag{9}$$

Substituting and solving yields

$$\underline{x} = (1-\theta)^{-1}\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)^{\alpha-1} \tag{10}$$

3.6. Equilibrium

The model's equilibrium is characterized by a wage which clears the labor market. That is to say, labor demanded by the formal sector plus the labor demanded by the informal sector equals the labor supplied by individuals who choose not to be entrepreneurs.

$$\int_{\overline{x}}^{1} n_F^* dx + \int_{\underline{x}}^{\overline{x}} n_I^* dx = \underline{x} \tag{11}$$

I can't solve this in closed form, or even sign the various derivatives associated with changing the parameters[2].

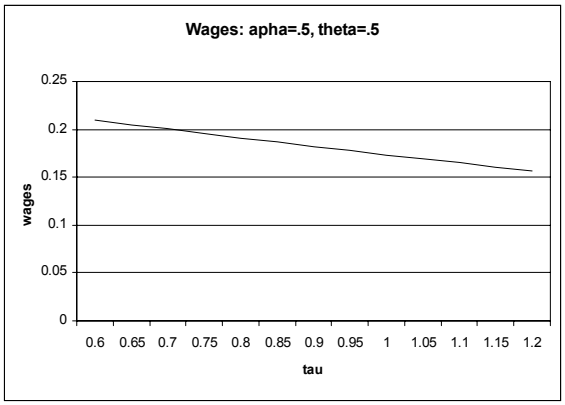
However, the model can be simulated numerically.

[2]One aspect of the model which can be deduced analytically is that productivity between different formal firms is constant regardless of size. The same is true for productivity across informal firms. Tybout (2000) finds empirical evidence for constant productivity across firms of different sizes.

3.7. Simulation Results

I examine different combinations of values for α and θ. For each combination, I vary τ from the level at which no informal firms exist (*i.e.,* rent seeking is so low that all entrepreneurs choose to operate in the formal sector) to the level at which no formal firms exist (*i.e.,* rent seeking is so high that all entrepreneurs choose to operate in the informal sector.) Over the whole range of both parameters, the effect of increased rent seeking does not vary. As demonstrated in Figure 1., as τ increases, wages are driven down.

Figure 1.[3]

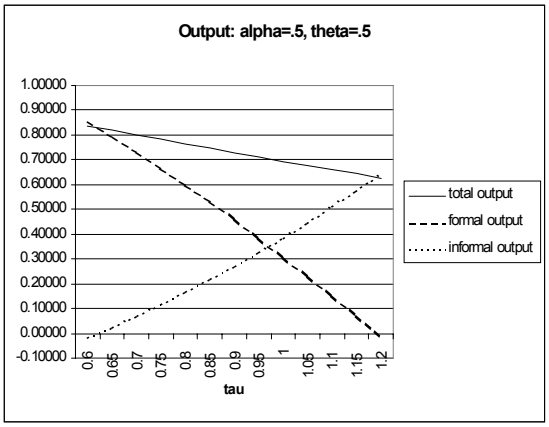


Entrepreneurs operating at the margin of the formal sector exit into the informal sector widening the gap between the largest informal firm and the smallest

[3]The scales on the axes for all the graphs are endogenous to the model. They do not correspond to real-world values and should be interpreted as guides to relative changes in magnitudes.

formal firm. As they do so, their productivity jumps down, as does the scale at which they operate. Formal sector output falls, informal sector output rises and overall output falls, as demonstrated in Figure 2.

Figure 2.



As wages fall, those entrepreneurs remaining in the formal sector hire more workers, thereby increasing the size of their firms. The informal sector grows from two directions. On the one hand, entrepreneurs exiting the formal sector start firms in the informal sector. On the other hand, lower wages entice the most entrepreneurially skilled workers to start their own informal firms.

This process is illustrated in Figures 3a and 3b., which illustrate the effect on firm size distribution from increasing the level of bureaucratic rent seeking τ.

Figure 3a.

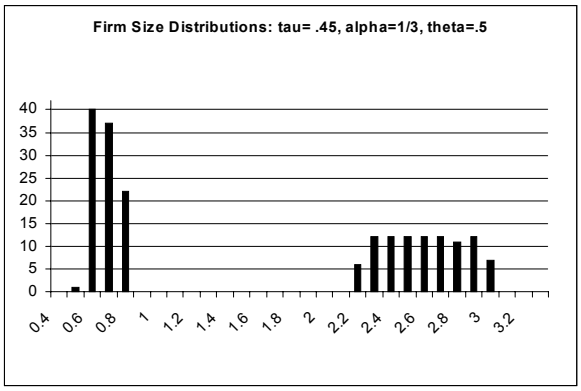
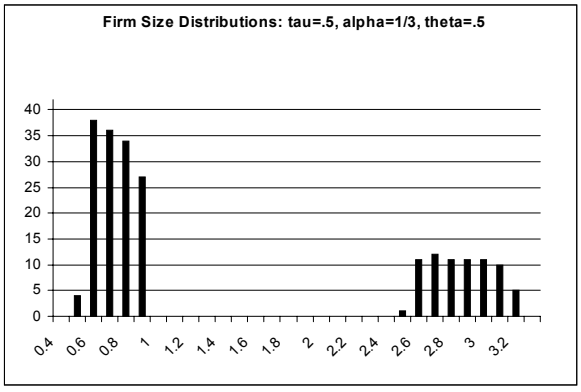


Figure3b.



3.8. Empirical Implications

Obviously, we would not expect such a simple model to conform exactly to observed data. In particular, it would be unreasonable to expect there to be a total absence of mid-size firms, as the model predicts. However, a number of broader testable implications can be formulated:

1) Both the output of the informal sector and the number of firms in the informal sector increases as τ increases.

2) Wages fall as τ increases.

3) Total output falls as τ increases.

4) The number of formal firms decrease as τ increases.

5) The average size of formal firms increase at τ increases.

6) The ratio of large formal firms to small formal firms increases as τ increases.

The first three implications are consistent with previous empirical work relating the overall size of the informal sector and per capita income to government rent seeking. I will concentrate on the last three implications since they relate specifically to the distribution of firms by size.

4. Empirical Results

To test the model I turn to two sources. As a measure of bureaucratic rent-seeking and red tape- which in the model are embodied in the parameter τ- I use data from Djankov, et al. (2000), which constructs measures of bureaucratic obstacles to firm creation. The authors gather data from 75 countries on the number of procedures, official time and official cost necessary for a new firm to operate legally. These measures fit very well with the sort of bureaucratic obstacles and rent seeking described by de Soto. The number of bureaucratic procedures is positively correlated with corruption, as measured by the Transparency International Index, and with the proportion of informal employment from Schneider (2000). I will

use the number of procedures as my primary measure of τ, although time and cost yield similar results.

Cross-country data on the distribution of firm sizes is very limited.. There are many studies of firm behavior *conditional* on firm size, but little data on the actual distribution of firms. The only extensive cross-country data set on firm size distributions is the United Nations Industrial Development Organization (UNIDO) Industrial Statistics Size-Distribution (ISSZ) Database. The data cover the manufacturing sector in 45 countries and are drawn from the national statistics offices of non-OECD countries and from the OECD for member countries. The variables of interest for this paper are the number of establishments by size class in terms of number of employees and the number of total employees by firm size class.

UNIDO attempts to find data conforming to the standard classification bins for firm size by number of persons employed set forth by the UN: 4-9 persons; 10-19 persons; 20-49 persons; 50-99 persons; 100-199 persons; 200-499 persons; 500-999 persons; and 1000 or more persons. Within the manufacturing sector, the data are broken down into 28 manufacturing branches at the 3-digit level of International Standard Industrial Classification (ISIC), Revision 2.

While this is the most extensive data set available, it is nonetheless problematical for my purposes. It only covers manufacturing; much of informal economic activity occurs in the service sector. The data are from a range of different years for different countries. I use the latest year reported for each country. Most observations are from the 1980's and 1990's, but the earliest stretch back to 1971. The

ISIC categories reported vary from country to country and, within given countries, from year to year. Many countries do not use the standard UN bin sizes and very few report observations for all of the standard bin sizes. Bin sizes vary by year and by ISIC code for the same country.

By way of example, consider two typical observations from the database, one of Japanese beverage producers and the other of Argentinian textile manufacturers.

Table 1. Beverage (ISIC=313) Manufacturing Establishments in Japan, 1990

Range of Size by # of Employees	# of Establishments	Total # of Employees
4-9	1292	8399
10-19	781	10804
20-49	553	16384
50-99	193	13636
100-199	102	14175
200-499	59	17329
500-999	7	4502
1000+	1	1389

Table 2. Textile (ISIC=321) Manufacturing Establishments in Argentina, 1993

Range of Size by # of Employees	# of Establishments	Total # of Employees
1-5	1520	1262
6-10	394	2252
11-25	442	6544
26-50	229	7747
51-150	185	14897
151-250	39	7387
251-400	20	6547
401+	11	7489

While Japan follows the UN standard bin sizes, Argentina does not.

It is also important to note that the data covers establishments instead of firms. An establishment is defined as, "A unit that engages, under a single ownership or control, in one, or predominantly one, kind of activity at a single location; for example, workshop or factory." (ISSZ documentation) By contrast, a firm or enterprise may consist of a number of different establishments. While it would better to have data on firms/enterprises, there is little reason to suppose that establishment size would not reflect the effects predicted by the model. Indeed, an entrepreneur might have two smaller workshops instead of one larger one precisely to remain in the informal sector.

I argue that the data cover **only formal firms** or, less strictly, that informal firms are observed much less frequently than formal firms. I will defend this assertion in section 4.2.

Population and percapita GDP data are taken from the Penn World Tables, using the year corresponding to the year of the observation in the ISSZ Database for each country.

4.1. Aggregate Regressions

Table 3. shows the regression of the natural logarithm of the number and average size of formal sector establishments on the log of the number of procedures, the log of per capita GDP, the log of population and dummy variables for each ISIC code. The coefficients for the ISIC codes are not reported. The coefficients for the log number of procedures are of the opposite sign predicted by the model. They are significant using standard errors, but are not significant using errors clustered by

country.

Table 3. Number and Average Size of Establishments with At Least 10 Employees.

	Log # of Est.	Log Ave. Size
Log # of Procedures	0.249	-0.235
	(1.24)	(-1.13)
Log Per Capita GDP	1.207	-0.068
	(8.89)	(-0.65)
Log Population	1.810	0.799
	(7.92)	(2.11)
R^2	0.795	0.445
Adjusted R^2	0.780	0.404
# of Observations	798	798
# of Countries	36	36

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 4. presents regression results testing whether small formal establishments are pushed out of the formal sector by high levels of rent seeking. To measure this potential effect, I construct a measure of the distribution of formal sector establishments. The left hand side of the specification is,

$$LHS = \frac{\text{\# of Establishments with 50+ Employees}}{\text{\# of Establishments with 10+ Employees}} \tag{12}$$

If the size of the smallest formal establishments fall in the range of 10-49 employees and small formal establishments are driven out of the formal sector by rent seeking, then when this ratio is regressed on the log number of procedures, the sign of that coefficient should be positive. In fact, it is negative, though not significantly using country clustered errors

Table 4. Establishments with 50+ Employees as a Percentage of All Establishments with At Least 10 Employees

Log # of Procedures	-0.060
	(-1.73)
Log Per Capita GDP	-0.078
	(-3.33)
Log Population	0.032
	(0.73)
R^2	0.510
Adjusted R^2	0.459
# of Observations	530
# of Countries	24

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

The comparison of establishments with 10-49 employees to establishments with 50 or more employees is chosen for two reasons. First, the literature on informal firms suggests that most are smaller 10 employees. Similarly, firms with greater than 50 employees can be presumed to be operating in the formal sector. If relatively small formal sector firms are disappearing due to rent seeking, they should fall somewhere in this range.

Second, observations containing the UN standard bin sizes of 10-49 are relatively common in the dataset. Observations not containing this bin size are not included in this regression.

Table 5. presents two more regressions using the same methodology, but which examine other bin sizes: 20-49 vs. 50+ and 20-99 vs. 100+. Once again, the sign of the coefficients for the log of the number of procedures is negative, the opposite of the model's predictions. Both are significant at the 5% level with country clustered errors

Table 5. Larger Establishments as a Percentage of
All Establishments with At Least 20 Employees

	% of Est. with 50+ Employees	% of Est. with 100+ Employees
Log # of Procedures	-0.063	-0.058
	(-2.05)	(-2.07)
Log Per Capita GDP	-0.070	-0.059
	(-3.46)	(-2.88)
Log Population	0.038	0.046
	(0.26)	(1.42)
R^2	0.454	0.4866
Adjusted R^2	0.390	0.4211
# of Observations	490	443
# of Countries	23	21

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

4.2. Do the Data Include Only Formal Sector Establishments?

I offer three arguments to support my claim that the establishments considered above are all or mostly in the formal sector. First, all of the data in the ISSZ Database are from government agencies. By the very nature of the informal sector, government enumerators will face great difficulty in collecting data on informal establishments. Such establishments are not paying taxes and are operating without permits. They are actively trying not to call government attention to their operations. Even if government enumerators seek to count them, informal establishments will surely be under-counted at a much higher rate than formal sector establishments.

Second, even if informal firms are included in some of the observations, the model (along with the previous empirical work cited above) predicts that they

should all have few employees; most likely fewer than ten, certainly fewer than twenty. The regressions for the total number of establishments and the average number of establishments only consider establishments with 10 or more employees. The bin size regressions similarly examine only firms with more than 10 or more than 20 employees.

Third, I am able to split the sample into observation which may include some informal firms and those which most likely do not. UNIDO does not compile methodologies used by each country's national statistics office. However, there are two distinct approaches to gathering the data: industrial censuses and annual surveys.

> In general, the size-distribution data were outcomes of either industrial censuses or of annual industrial surveys. If the source is a census, the reference establishments are usually all establishments regardless of their legal status or size. On the other hand, if the source is an annual survey then the reference establishments are those registered at the national statistics office for the survey purpose (They are usually above a certain employment cut-off point, e.g., those with 5 or more persons engaged). Roughly speaking, one may assume that the cases where data for the size class with lower end as one person are included were derived from censuses and that the other cases were derived from annual surveys.[4]

[4] Personal communication from UNIDO Chief Statistian Tetsuo Yamada, September 12, 2002

If an establishment is registered with a national statistics office then by my definition it is in the formal sector. By splitting the sample into observations which have one as lower bound of their bin sizes and observations in which the lower bound of their bin size is greater than one, I am able to evaluate whether the two methodologies yield markedly different results, suggesting perhaps that informal firms appearing in the industrial censuses in significant numbers. As shown in Tables 6, 7 and 8, without exception, the sign of the coefficient of the log of the number of procedures does not change when the sample is split.

Table 6. Number and Average Size of Establishments
with At Least 10 Employees, Smallest Bin Size (SBS)=1,≠1

	Log # of Est.		Log Ave. Size	
	SBS=1	$SBS \neq 1$	SBS=1	$SBS \neq 1$
Log # of Procedures	0.232	0.320	-0.142	-0.389
	(0.72)	1.25	(-1.05)	(-1.13)
Log Per Capita GDP	1.209	1.234	-0.094	-0.078
	(5.30)	(6.61)	(-0.90)	(-0.46)
Log Population	2.085	1.543	0.455	1.408
	(7.87)	(3.72)	(6.20)	(1.70)
R^2	0.820	0.793	0.571	0.489
Adjusted R^2	0.792	0.764	0.505	0.420
# of Observations	381	417	381	417
# of Countries	18	20	18	20

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 7. Establishments with 50+ Employees as a Percentage of All Establishments with At Least 10 Employees, Smallest Bin Size $(SBS) = 1, \neq 1$

	$SBS = 1$	$SBS \neq 1$
Log # of Procedures	-0.073	-0.035
	(-1.53)	(-0.74)
Log Per Capita GDP	-0.069	-0.081
	(-1.71)	(-2.60)
Log Population	0.065	-0.019
	(1.82)	(-0.19)
R^2	0.541	0.554
Adjusted R^2	0.449	0.461
# of Observations	257	273
# of Countries	12	13

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 8. Larger Establishments as a Percentage of All Establishments with At Least 20 Employees, Smallest Bin Size $(SBS) = 1, \neq 1$

	% of Est. with 50+ Employees		% of Est. with 100+ Employees	
	$SBS = 1$	$SBS \neq 1$	$SBS = 1$	$SBS \neq 1$
Log # of Procedures	-0.074	-0.055	-0.045	-0.054
	(-1.65)	(-1.07)	(-1.58)	(-0.99)
Log Per Capita GDP	-0.062	-0.075	-0.012	-0.071
	(-1.82)	(-2.71)	(-1.03)	(-2.80)
Log Population	0.054	0.041	0.070	0.031
	(1.78)	(0.40)	(4.13)	(0.33)
R^2	0.493	0.493	0.628	0.507
Adjusted R^2	0.385	0.372	0.530	0.371
# of Observations	240	250	226	217
# of Countries	11	13	11	11

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

While the sign on the coefficient of the log of the number of procedures doesn't

change, the magnitude of the coefficient is larger for the industrial census observations in four of the five specifications. One possible explanation of this could be that informal firms are indeed contaminating these observations to some degree. In order to examine degree of difference in the coefficients, I conduct Wald tests with the null hypothesis that both sets of coefficients are equal. For each specification, the null is not rejected. Table 9. reports these results.

Table 9. Wald Tests, Null: (Coefficient of Log # of Procedures for Observations with Smallest Bin Size = 1) = (Coefficient of Log # of Procedures for Observations with Smallest Bin Size≠1)

	F	Deg. of Free.	p
Number of Establishments with At Least 10 Employees.	0.05	35	0.832
Average Size of Establishments with At Least 10 Employees.	0.46	35	0.504
Establishments with 50+ Employees as a Percentage of All Establishments with At Least 10 Employees	0.64	22	0.431
Establishments with 50+ Employees as a Percentage of All Establishments with At Least 20 Employees	0.21	23	0.930
Establishments with 100+ Employees as a Percentage of All Establishments with At Least 20 Employees	0.00	20	0.963

Variance/covariance matrix estimated using country clusters.

4.3. OECD vs. Non-OECD

Splitting the sample into OECD and Non-OECD does yield more change in the coefficients. In two of the five specifications, the sign on the coefficient for the log of the number of procedures changes. However, the coefficients for the specifications involving establishments of twenty or more employees do not change.

Indeed, they are almost identical. One interpretation of this is that the smallest formal establishments are all larger than 20 employees and that the specifications including establishments with 10-19 employees are including informal firms in the Non-OECD countries. Tables 10, 11 and 12 report these results.

Table 10. Number and Average Size of Establishments with At Least 10 Employees, OECD/Non-OECD.

	Log # of Est.		Log Ave. Size	
	OECD	Non-OECD	OECD	Non-OECD
Log # of Procedures	0.219	0.402	0.166	-0.360
	(0.85)	(0.87)	(1.14)	(-0.94)
Log Per Capita GDP	0.976	1.228	0.246	-0.106
	(2.44)	(5.46)	0.87	-0.70
Log Population	1.839	2.001	0.049	1.207
	(6.28)	(4.56)	(0.29)	(2.07)
R^2	0.823	0.736	0.618	0.521
Adjusted R^2	0.797	0.701	0.560	0.459
# of Observations	354	444	354	444
# of Countries	15	21	15	21

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 11. Establishments with 50+ Employees as a Percentage of All Establishments with At Least 10 Employees, OECD/Non-OECD

	OECD	Non-OECD
Log # of Procedures	-0.041	0.028
	(-0.59)	(0.38)
Log Per Capita GDP	0.012	-0.099
	(0.08)	(-4.33)
Log Population	-0.004	-0.012
	(-0.05)	(-0.16)
R^2	0.522	0.587
Adjusted R^2	0.434	0.494
# of Observations	258	272
# of Countries	11	13

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 12. Larger Establishments as a Percentage of All Establishments with At Least 20 Employees, OECD/Non-OECD

	% of Est. with 50+ Employees		% of Est. with 100+ Employees	
	OECD	Non-OECD	OECD	Non-OECD
Log # of Procedures	-0.038	-0.032	-0.042	-0.040
	(-0.98)	(-0.35)	(-0.82)	(-0.60)
Log Per Capita GDP	0.036	-0.110	-0.015	-0.087
	(0.35)	(-4.47)	(-0.13)	(-3.42)
Log Population	-0.008	0.028	0.012	0.046
	(-0.14)	(0.43)	(0.21)	(0.81)
R^2	0.505	0.569	0.544	0.579
Adjusted R^2	0.422	0.425	0.462	0.420
# of Observations	289	201	263	180
# of Countries	13	10	11	10

Country clustered t-statistics are in parentheses. ISIC industry dummies are not reported.

Table 13. reports Wald tests with the null hypothesis that both sets of coefficients for the log of the number of procedures are equal. For each specification,

the null is not rejected.

Table 13. Wald Tests, Null: (Coefficient of Log # of Procedures for OECD Countries) = (Coefficient of Log # of Procedures for Non-OECD Countries)

	F	Deg. of Free.	p
Number of Establishments with At Least 10 Employees.	0.12	35	0.728
Average Size of Establishments with At Least 10 Employees.	1.69	35	0.203
Establishments with 50+ Employees as a Percentage of All Establishments with At Least 10 Employees	1.79	23	0.166
Establishments with 50+ Employees as a Percentage of All Establishments with At Least 20 Employees	0.01	22	0.943
Establishments with 100+ Employees as a Percentage of All Establishments with At Least 20 Employees	1.74	23	0.141

Variance/covariance matrix estimated using country clusters.

5. Conclusion

Having developed a formal model of the de Soto hypothesis and drawn predictions of firm size distributions form that model, I have found those predictions to not hold empirically. Why?

The simplest explanation is that the data are wrong. The ISSZ data set is far from perfect. The variations in observation years, manufacturing subsectors covered and bin sizes are dramatic. The lack of complete methodological descriptions and, presumably, significant variation in the methodology used by different countries can not be ignored.

It may also be the case that manufacturing is not the place to look for confir-

mation of the de Soto hypothesis. Perhaps data on the service sector would bear out his predictions. Perhaps there is some other effect of rent seeking and bureaucratic obstacles that leads to smaller formal firms and that this effect swamps the opposite tendencies implicit in de Soto. Fisman and Svensson (2000) find just such an effect on Ugandan firms.

Finally, it may be the case that the de Soto is wrong. However, such a conclusion is surely premature. Instead, the results presented here should be taken as an impetus for further research on firm size and informality. Most importantly, governments and international agencies interested in small and medium sized firms should put greater emphasis on developing better data on firm size distributions.

References

[1] Yoshiaki Azuma and Herschel I. Grossman. A theory of the informal sector. *Brown University Working Paper*, 2002.

[2] Omar D. Bello. The informal sector, productivity shocks and distortions in the labor market. *Mimeo*, 1999.

[3] Jay Pil Choi and Marcel Thum. Corruption and the shadow economy. *CESifo Working Paper*, 633, January 2002.

[4] Simeon Djankov, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Schleifer.

[5] Raymond Fisman and Jakob Svensson. Are corruption and taxation really harmful to growth? firm level evidence. *Mimeo*, May 2000.

[6] Bernard Fortin, Nicolas Marceau, and Luc Savard. Taxation, wage controls and the informal sector. *Jornal of Public Economics*, 66:293–312, 97.

[7] Eric Friedman, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton. Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of Public Economics*, 76:459–493, 2000.

[8] Takao Fukuchi. A simulation analysis of the urban informal sector. *The Developing Economies*, 36(3):225–56, September 1998.

[9] Robert Hall and Charles Jones. Why do some countries produce so much more output per capita than others. *Quarterly Journal of Economics*, 114(1):83–116, February 1999.

[10] John R. Harris and Michael P. Todaro. Migration, unemployment and development: A two-sector analysis. *American Economic Review*, 60(1):126–142, January 1970.

[11] Peter Klenow and Andres Rodriguez-Clair. The neoclassical revival in growth economics: Has it gone too far? *NBER Macro Annual*, 1997.

[12] Maria Lacko. Hidden economy in east european countries in international comparison. *International Institute for Applied Systems Analysis Working Paper*, 1996.

[13] Norman V. Loayza. 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, 1996.

[14] Robert E. Lucas, Jr. On the size and distribution of business firms. *Bell Journal of Economics*, 9(2):508–523, September 1978.

[15] William F. Maloney. Are ldc labor markets dualistic. *World Bank*, 1998.

[16] N. Gregory Mankiw, David Romer, and David Weil. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2):407–437, May 1992.

[17] Douglas Marcouiller and Leslie Young. The black hole of graft: The predatory state and the informal economy. *The American Economic Review*, 85(3):630–646, June 1995.

[18] Paolo Mauro. Corruption and growth. *Quarterly Journal of Economics*, 110(3):681–712, August 95.

[19] James E. Rauch. Modelling the informal sector formally. *Jounal of Development Economics*, 35:33–47, 1991.

[20] Pierre-Daniel G. Sarte. Informality and rent-seeking bureaucracies in a model long-run growth. *Journal of Monetary Economics*, 46:173–197, 2000.

[21] Freidrich Schneider and Dominik H. Enste. Shadow economies: Size, causes and consequences. *Journal of Economic Literature*, 38:77–131, March 2000.

[22] Friedrich Schneider. The shadow economies of western europe. *Journal of the Institute of Economic Affairs*, 17(3):42–48, 1997.

[23] Hernando de Soto. *The Other Path: The Invisible Revolution in the Third World*. Harper Row, New York, 1989.

[24] James Tybout. Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature*, 38(1):11–44, March 2000.

[25] Colin C. Williams and Jan Windebank. Black market work in the european community: Peripheral work for peripheral localities. *Internation Journal of Urban and Regional Research*, 19(1):23–39, 1995.

[26] Gustavo Yamada. Urban informal employment and self-employment in developing countries: Theory and evidence. *Economic Development and Cultural Change*, pages 289–314, 1996.