Income Distribution and the Size of the Informal Sector*

Diego Winkelried

St John's College, University of Cambridge

CB2 1TP, Cambridge, UK

dw295@cam.ac.uk

This version: October 25, 2005

Abstract

This paper studies the role of income distribution as a determinant of the size of the informal sector in an economy by relying on a channel whereby inequality affects the behaviour of aggregate demand and thus influences the incentives a firm has to become informal. It is further postulated that income distribution affects the response of the informal sector to different fiscal policies, either demand or supply-orientated. The main findings are that high inequality leads to a large informal sector, and that redistribution towards the middle class decreases the size of the informal sector and increases the capacity of fiscal instruments to reduce informality. Empirical evidence for Mexican cities is provided.

Keywords : Income distribution, market size, informal sector.
JEL Codes : D31, O11, O17.

*I would like to thank Professor Partha Dasgupta, Paul Castillo, Rafael De Hoyos, Carlos Montoro, Arina Nikandrova and seminar participants at the University of Cambridge and the Central Bank of Peru for helpful comments on early drafts of the paper. The opinions expressed herein and any remaining errors are mine. Financial support from the Cambridge Overseas Trust is gratefully acknowledged.

1 Introduction

The literature on the driving forces underlying the size of informal economy has mainly focused on the effects of government actions, notably taxation and regulation[1], and has reached the widespread conclusion that the existence of an informal sector is the result of the failure of political institutions to promote a working market economy[2]. Consistently with this view, Johnson *et al.* (1998) and Friedman *et al.* (2000) find that institutional traits (for instance, the extent of corruption or the strength of the rule of law) explain most of the cross-country variation of the available informality measures[3].

It is also well-established that when studying informality, firm size does matter. Large, capital-intensive firms are prone to operate in the formal sector, reaping the fruits of efficiency from legality and economies of scale, while small firms usually operate informally. They sacrifice efficiency but avoid taxation. In words of De Soto (1989, p. 153),

> "[*T*]he need to avoid detection forces informals to operate on a very small scale. They 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. [*S*]uch arrangements [...] prevent them from achieving efficient scales of production."

On the other hand, Levenson and Maloney (1998) provide a fresh, alternative interpretation to the emergence of an informal sector, in particular to the size choice of informals, without relying on the burden imposed by the government to the private sector. They state that small firms do not scale down to avoid paying taxes, but their limited investment needs and the narrow nature of their operations make stable property rights unimportant and the gains from civic participation flimsy. Naturally, since the benefits from participating in societal institutions grow larger as firms do, voluntary compliance and the will to being charged (i.e., taxed) for participation arise[4]. In other words, firms evolve from informality to formality as they grow to their long-run equilibrium size.

Reasoning along these lines, the structure of the market, especially the demand that firms face, is likely to be a determinant of the size of the informal sector as important as the

[1] See De Soto (1989) and Rauch (1991). See also Schneider and Enste (2000) for a review.

[2] Loayza (1996) provides a balance of costs and benefits of (in)formality. See also Schneider (2005).

[3] A striking finding of Friedman *et al.* (2000) is that higher tax rates are associated with a small informal sector. They argue that high tax rates increase tax revenues that would enable the government to finance a stronger legal environment and, consequentially, to reduce informality.

[4] An interesting conclusion from this analysis is that since it is not cost-effective for the government to monitor the smallest and typically least productive firms, the large informal sectors observed in developing countries may be optimal. See Asea (1996).

governmental burden imposed on business-making[5]. In fact, a firm deciding the sector in which to operate would compare the benefits from producing with scale economies and paying taxes against the profits of producing under a less efficient technology. If a demand expansion occurs, *ceteris paribus*, the benefits of formality become evident as it eases meeting the higher demand and generating the corresponding profits, leading to a further reduction in the costs of formality.

In this study, I argue that income distribution plays a role in the determination of the size of the informal sector as it shapes the way the market, particularly aggregate demand, behaves[6]. In this fashion, income distribution influences the environment in which a firm decides its size and its formality status. Furthermore, as a by-product of the analysis, I postulate that income distribution affects the response of the informal sector to different fiscal policies, either demand or supply-orientated. The literature is basically silent on these matters whose importance lies in the fact that they may add a new perspective in evaluating the effects of redistributive policies, especially in developing countries. Indeed, there are just a few previous works exploring the relationship between income distribution and the size of the informal economy, all of them based on different theoretical motivations from the one stressed here.

Rauch (1993) finds, in a model with rural-urban migration, that the informal sector share of the total labour force follows an inverted-U pattern with the level of urbanisation, linking in this way informality with the well-known Kuznets's hypothesis. During the middle stage of development, inequality increases as many poor move to the city and participate in the informal sector. Schneider and Enste (2000, ch. 7) argue that a large social welfare system aimed to flatten the income distribution should increase the size of the informal economy because of strong disincentives to work in the formal economy, although this heavily depends on the nature of the public transfers programs (whether it is targeted or redistributive).

Rosser *et al.* (2000, 2003), on the other hand, find a positive correlation between the Gini coefficient and the size of the informal share among transition economies. They argue that the detrimental effect of informality on public finances reduces the capacity that a government has to perform sound redistributive policies, whereas inequality encourages the desire a person may have to beat the system and to not comply with the prevailing regulations. Finally, Chong and Gradstein (2004) perform an exhaustive empirical cross-country analysis and find that higher inequality in conjunction with weak institutions increases the degree of informality. They suggest that when resources are up for grabs in

[5] According to recent findings in Antunes and Cavalcanti (2004), regulations costs are important in explaining the size of the informal sector in industralised economies, whereas market imperfections are far more important in developing countries.

[6] Benassi *et al.* (2002) find that an increase in the income concentration towards the middle class expands demand and increases its price elasticity. This idea, however, goes back to Robinson (1969, pp. 70–71).

the formal sector, poor individuals may find it beneficial to enter the informal sector where they are able to retain their production output fully. High inequality, exacerbated by low institutional quality, magnifies this effect.

The rest of the paper is organised as follows: section 2 presents a stylised model in which income distribution ultimately affects the decision of firms to become (in)formal. In section 3, I analyse the equilibrium properties of the model and derive empirical hypotheses to be tested in section 4, with data from Mexican cities. Section 5 concludes and suggests some avenues for future research.

2 A simple model

Next a simple model where income distribution determines demand[7] and firms decide whether to operate in the formal or the informal economy is outlined. The government collects taxes and returns them to society either as a productive public good for its use by formal firms or as transfers to the poor. For expositional convenience, the institutional quality of the model economy is given and it is assumed that firms in both sectors exercise full property rights over their production. Moreover, I refrain from modeling distracting issues for the purposes of this study such as corruption, imperfect enforceability, and credit market imperfections.

2.1 Production

There is a continuum of productive sectors. Each one consists of just one competitive firm which produces a homogenous good using labour, which is also homogenous, as the unique input. The input requirement for any sector $z \in [0, \infty)$ to produce a unit of output is $\alpha > 1$ and labour is paid a wage w.

2.1.1 Types and prices

In each sector, the firm can be either of two types: informal or formal. An informal firm has profits $\pi_z^I = (p_z - \alpha w)Q_z$, where p_z and Q_z are, respectively, the price and quantity produced of good z. Competition implies that in equilibrium $p_z = p = \alpha w$ and $\pi_z^I = 0$.

Alternatively, a firm can belong to the formal sector. If so, it must pay a fixed entry fee τ to the government, which can be interpreted as a tax rate, and it receives a benefit in exchange: its unit labour requirement reduces to $\alpha(1 - \rho) < \alpha$, where $\rho \in (0, 1)$ is a

[7] The demand side of the model is based on Dasgupta and Ray (1986), Murphy *et al.* (1989) and Matsuyama (2000), which analyse different aspects and consequences of income inequality.

productivity-enhancing factor that arises from the access to productive public services[8]. Hence, the profits of a formal firm are $\pi_z^F = (p_z - \alpha(1-\rho)w)Q_z - \tau$[9].

Since goods are homogenous, a formal firm may compete with informals. Charging a price higher than p implies zero sales for the former while charging a lower price rules out informal firms. To maximise profits, the price charged will be infinitesimally below p, so, in equilibrium, a formal firm charges what it would be charged by an informal $p_z = p = \alpha w$. Thus, $p = \alpha w$ for every sector, regardless of the size of the informal sector and of the type of the firm[10].

2.1.2 Choosing (in)formality

Consider a sector q in which the firm is indifferent between being formal or informal, i.e. $\pi_q^F = \pi_q^I = 0$. If N_q denotes the demand for good q, the zero profits condition implies that

$$N_q = \frac{\tau}{\alpha w \rho} \tag{1}$$

For given τ, ρ and w, the firm's decision to be in the formal sector depends on the demand it faces. From (1) it can be seen that if a firm z has a demand $Q_z < N_q$ then it will be informal as $\pi_z^F < 0 = \pi_z^I$; contrarily, if $Q_z > N_q$, the firm benefits from formality since $\pi_z^F > 0 = \pi_z^I$. Thus, N_q is the break-even demand level (or *size*) of a formal firm. As such, it is easy to see that a rise in N_q would make it harder for firms to enter the formal sector, as the minimum demand requirement to make formality worthy increases.

2.2 Households

There is a continuum of length one of households, each inhabited by one individual, which are identical in all aspects except in their income from rents. As ownership of the productive firms is assumed to be random, so is income.

2.2.1 Preferences and choices

A typical household consumes only a unit of each good z and has a utility $U = \int_0^\infty \beta_z x_z \, \mathrm{d}z$ where $x_z = 1$ if good z is consumed and $x_z = 0$ if not. I assume that β_z *is decreasing in* z.

[8] See Johnson *et al.* (1997) and Friedman *et al.* (2000) for alternative public finance mechanisms that relate tax revenue to the provision of public goods.

[9] It is important to note that the public service is non-rival and non-excludable and hence is not subject to congestion. As concluded in Barro and Sala-i-Martin (1992), in this case lump sum taxation is superior for the productive efficiency of public services to distortionary taxation.

[10] If informal firms are interpreted as incumbents and a formal firm as an entrant, we could rely on Murphy *et al.* (1989) to show that charging $p = \alpha w$ is a Nash equilibrium in such an entry game.

This implies that the individual consumes good z as long as she has consumed all the other goods with indices less than z. In other words, goods are ordered in terms of priority[11].

If the household's income is y, then its budget constraint is $\int_0^\infty p_z x_z \, \mathrm{d}z \leq y$ and the solution of the utility maximisation problem is straightforward: this household chooses $x_z = 1$ for $z \in [0, z^*]$ and $x_z = 0$ for $z \in (z^*, \infty)$ where z^* is given by

$$y = \int_0^{z^*} p_z \, \mathrm{d}z = \alpha w z^* \tag{2}$$

Note that since there is a direct mapping from U to z^*, the numbers of goods consumed z^* and income y are themselves indirect utility measures. Thus, an increase in utility is achieved only by augmenting the number (variety) of goods the household purchases.

Preferences are non-homothetic (income changes the marginal utility over goods) and lead to demand complementarity (in the sense of Hicks-Allen) from lower-indexed goods to higher-indexed goods, but not the other way around. Thus, an increase in income or a reduction in p increases the demand for higher-indexed goods without affecting the demand for lower-indexed goods. To be more precise, the marginal propensity to consume a good with a higher index than z^* is one, whereas it is zero for any good $z < z^*$.

2.2.2 Income distribution and poverty

Income is composite by wages and rents. Every household supplies inelastically a unit of labour and receives a wage w in exchange. It also owns a share $\theta \in [0; [\theta_p, \infty))$ of all the profits in the economy π (defined below), so its income is $y = w + \theta\pi$.

The shares are given randomly. Let $\mathcal{R}(\theta)$ be the number of people whose share is less than or equal to θ. The number of shareholders is $N_p = 1 - \mathcal{R}(\theta_p)$, so there are $\mathcal{R}(\theta_p)$ people with income $y = w$ who cannot purchase any good traded in the market since $p = \alpha w > w$. I will refer to the subset $\mathcal{R}(\theta_p)$ as the "poor."

2.2.3 Aggregate demand

Given the non-homotheticity of preferences, income distribution determines aggregate demand[12]. Households with income greater than $y_z = w + \theta_z \pi$ purchase one unit of good z. Since each household buys only one unit of each good she can afford, the aggregate demand

[11] An alternative setup that also leads to a well-defined preference ordering over goods is to set $\beta_z = \beta$ and let the input requirement vary across sectors. The priorities will be given by the ranking among the prices of the goods. See Matsuyama (2000) for further details.

[12] It is well-known that with homothetic preferences only the mean of income matters in the determination of demand. The results of the present model depend crucially on this non-homotheticity assumption. If preferences were homothetic and no further restriction were imposed on household income (for instance, a liquidity constraint), then the multiplier effect explained below would not arise.

of good z is

$$N_z = 1 - \mathcal{R}(\theta_z) \tag{3}$$

Goods at the lower end of the spectrum are consumed by almost all households. As their income levels go up, the households expand their range of consumption by adding higher-indexed goods to their basket. Hence, rich households consume the same goods as the poor plus some more, implying that N_z is decreasing in z. Therefore, following the firms' decision surrounding equation (1), firms in sectors $z \in [0, q]$ will be formal whereas firms in $z \in (q, \infty)$ prefer informality.

2.3 Fiscal balance

For the moment, the role of the government in the model is to collect taxes and return them back to society in the form of the productivity factor ρ. An alternative use of fiscal resources in the form of a transfers program is analysed in section 3.3.

Let g be the fiscal expenditure designated to the provision of the public service. As tax receipts come from formal firms only, the fiscal balance identity is

$$g = \int_0^q \tau \, \mathrm{d}z = \tau q \tag{4}$$

The public service is produced with the following technology:

$$\rho = \rho(g) \quad \text{where} \quad 0 < \lim_{g \to 0} \rho < \lim_{g \to \infty} \rho < 1 \tag{5}$$

The elasticity of ρ with respect to the public expenditure $\varepsilon_g = (\mathrm{d}\rho / \mathrm{d}g)(g/\rho) \geq 0$ is assumed to be small[13].

2.4 Equilibrium

The expenditure of a household that consumes goods in the interval $[0, q]$ is $pq = \alpha w q$ whereas the income that allows it to purchase these goods is $y_q = w + \theta_q \pi$, hence

$$\alpha w q = w + \theta_q \pi \tag{6}$$

Now, equilibrium can be found upon aggregation of the individual expenditures.

[13] The appendix, equation (A15), shows what is meant by "small ε_g" in terms of the various parameters of the model.

2.4.1 Formal sector

From the above discussion, note that consumers with $\theta \in [\theta_p, \theta_q]$ spend all their income in goods sold in the formal sector. On the other hand, consumers with $\theta \in (\theta_q, \infty)$ only spend the share θ_q of their rents in such goods. Hence, the total sales of the formal sector equal the aggregate spending in goods $z \in [0, q]$,

$$S^F = \alpha w Q^F = \int_{\theta_p}^{\theta_q} \min\{w + \theta\pi, w + \theta_q\pi\} \, \mathrm{d}\mathcal{R}(\theta) = wN^p + \pi\left(\Theta_q + \theta_q N_q\right) \tag{7}$$

where Q^F is the aggregate production in the formal sector and $\Theta_q \equiv \int_{\theta_p}^{\theta_q} \theta \, \mathrm{d}\mathcal{R}(\theta)$ is the share of profits in the hands of the people belonging to the "middle class."

It is useful to compare (7) with the overall sales (in both sectors):

$$S = \alpha w(Q^F + Q^I) = \int_{\theta_p}^{\infty} (w + \theta\pi) \, \mathrm{d}\mathcal{R}(\theta) = wN_p + \pi\Theta \tag{8}$$

where $\Theta \equiv \int_{\theta_p}^{\infty} \theta \, \mathrm{d}\mathcal{R}(\theta)$ and Q^I is the supply of goods from the informal sector. It can be seen that while total sales are determined by the income of all shareholders, the sales of the formal sector depend on the income of the middle class and a fraction of the income of the "upper class" (the subset with $\theta > \theta_q$). This fact follows from the demand structure of the model: as the lower-indexed goods are at the top of the preference ranking, aggregate demand will be high and hence the benefits from formality are sizeable in those sectors. Thus, a wealthier individual can afford buying higher-indexed goods so that she contributes with the sales of low demand firms[14].

Aggregate profits are $\pi = \alpha w \rho Q^F - \tau q$. Rearranging using (1), (6) and (7) leads to

$$\pi = \left(\frac{1}{1 - \rho\Theta_q}\right) \rho w (N_p - N_q) \tag{9}$$

Remarkably, the sales to the upper class allow formal firms to face the entry costs τ[15]. The remaining sales, those aimed at the middle class, give formal firms profits that are eventually redistributed among households. When an individual enters the middle class

[14] *A caveat.* It may seem counterintuitive that the upper class purchases informal goods whereas the poor and the middle class purchase only formal goods. This result follows from the fact that prices (and qualities) among the traded goods are equal, so demand is determined by the number of customers. If prices were allowed to change, as suggested in footnote 11, then demand would be determined by the expenditure (or income) of the customers, with the intuitive result that those sectors with higher [lower] *value of sales* (i.e. revenues) would be [in]formal. This also would create a (positive) correlation between the costumers' wealth and sales, leading to informal goods being purchased by the poorest individuals. Allowing for prices or qualities between formal and informal goods to differ, would only increase the mathematical difficulty of the setup, without changing the insights of the model developed here.

[15] This is captured in the last term of equation (7), $\theta_q \pi N_q$. It is easy to show that $\rho\theta_q\pi N_q = \tau q$.

and purchases a good in the formal sector, profits increase by $\rho w(N_p - N_q)$ where ρ is the profit rate of formal firms. Then, a fraction of Θ_q of the generated profits is redistributed to the middle class and comes back as demand for further formal goods, leading to an extra increase in profits of $\rho\Theta_q$. As this process continues indeterminately, a multiplier effect arises with $(1 - \rho\Theta_q)^{-1} > 1$ being the multiplier.

2.4.2 Labour market

The labour market equilibrium condition is $L^F + L^I = 1$ or $\alpha(1-\rho)Q^F + \alpha Q^I = 1$. Provided that labour supply is fixed, that each household supplies inelastically a unit of labour and that the same labour input is demanded by the formal and informal sectors, there exists a positive wage rate that clears the labour market. From now on, the wage is treated as the *numeraire*, $w = 1$.

It is worth noting that as $L^I = \alpha Q^I$, an increase in L^I is associated with a proportional increase in the sales of the informal sector and vice versa. This tight relationship does not hold necessarily in the formal sector, $L^F = \alpha(1 - \rho)Q^F$, where an increase in L^F may be associated with a higher input requirement, i.e. a smaller ρ, even when sales are kept unchanged. Moreover, as the labour supply is constant, a change in L^F is compensated with a similar change (with the opposite sign) in L^I. Consequently, if the goods market in the formal sector is in equilibrium, so are the labour market (in both sectors) and the goods market in the informal sector. Thus, equations (1) and (9) alone define the goods and labour market equilibria.

2.4.3 General equilibrium

Apart from the goods and labour markets described, there is a third market that has not been modelled explicitly, which entails the transactions carried "outside the market" by the $\mathcal{R}(\theta_p)$ people defined as poor, with an aggregate income of $w\mathcal{R}(\theta_p) = \mathcal{R}(\theta_p)$. This market has many interpretations, for example, that of a subsistence good that is a substitute of the good traded in the model economy (though the latter is strictly preferred). By Walras's law, equilibrium is reached if both (1) and (9) are solved. Notice that we are implicitly assuming that the profits in this market are zero and, hence, do not contribute with aggregate wealth[16].

[16] Furthermore any feedback from this market to the economy modelled, for instance in productivity is ruled out. A richer framework where such a feedback is present is Dasgupta and Ray (1986).

2.5 Mechanisms

Given $\mathcal{R}(\theta)$, two opposite forces lead to equilibrium following the setting of the tax rate τ. At the firm level, a tax increase has a negative effect on the profits on the marginal firm q as the savings from contracting less labour are more than offset by the higher tax burden[17]. This increases informality and drops tax revenues and profits. At the aggregate level, on the contrary, for a given tax base q, a tax increase raises the multiplier and in turn aggregate profits. It follows a rise in income, demand, formality and tax revenues. The same reasoning applies for a tax reduction.

Both of these effects interact and the consequence is the presence of multiple equilibria, as in Johnson *et al.* (1997), i.e, the coexistence of "bad" high informality - low tax revenues - low demand equilibria with "good" low informality - high tax revenues - high demand equilibria. Which equilibrium the economy ends in will depend on initial conditions and, crucially, on income distribution (i.e. the degree of inequality)[18].

3 Analysis

Equations (1) and (9) (with $w = 1$) represent two curves in the (N_q, π) plane,

$$\begin{aligned} N_q &= \frac{\tau}{\alpha\rho} && (M\text{-curve}) \\ \pi &= \left(\frac{\rho}{1-\rho\Theta_q}\right)(N_p - N_q) && (F\text{-curve}) \end{aligned}$$

and constitute the toolbox for the comparative statics analysis of this section[19].

The F-curve relates profit generation with the presence of informal firms. This scheme has a negative slope since a higher value of N_q, *ceteris paribus*, expels some firms from the formal sector and reduces aggregate profits. Alternatively, a reduction of π decreases the income of shareholders, contracting the demand of all firms and making N_q relatively higher.

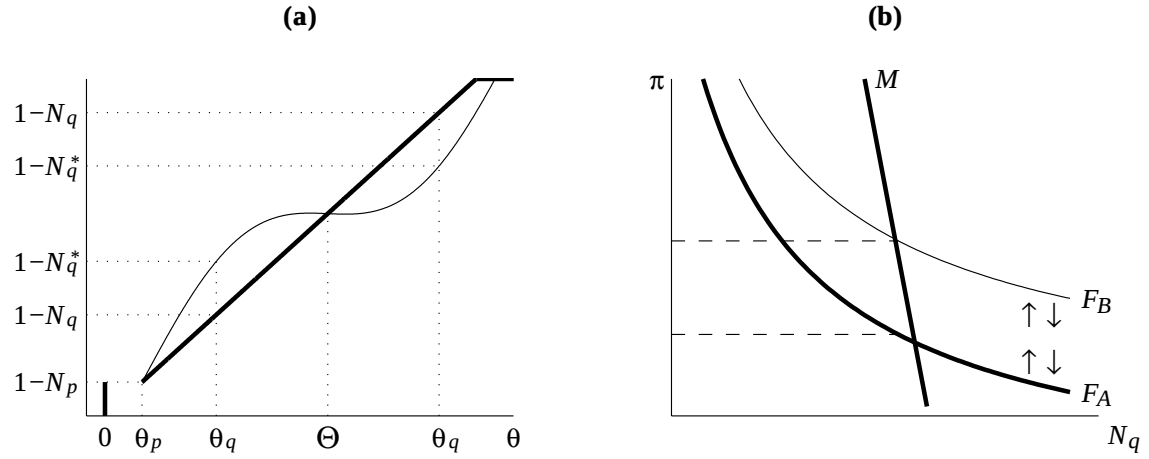
On the other hand, the points on the M-curve balance the incentives a firm has to become (in)formal. For given τ and α an increase in N_q should be accompanied by a proportional reduction in ρ so that ρN_q remains unchanged. The decrease in ρ comes from a fall in

[17] This follows directly from the definition of N_q in (1).

[18] These facts suggest that the "optimal" level of τ (and hence ρ) is an interior solution, as in Barro and Sala-i-Martin (1992) and Loayza (1996).

[19] See the appendix for the differential version of the model in which this section is based. The exposition on the following exercises may be interpreted as a dynamic response to shocks. Yet it is also useful to take them as cross-sectional comparisons of economies which is, coincidentally, the usual approach in empirical analysis of the informal sector.

Figure 1. Mean-preserving spread of $\mathcal{R}(\theta)$



the number of formal sectors q. This implies a reduction in profits and a downward-sloped M-curve. As the elasticity ε_g is assumed to be low, the M-curve is steeper than the F-curve.

3.1 Redistribution

To analyse the consequences of a redistribution of ownership on the informal sector, consider a mean-preserving spread of $\mathcal{R}(\theta)$ as shown in Figure 1(a), where the bold line represents some share distribution and the light line is its mean-preserving spread, $\mathcal{R}^*(\theta)$. A shift from $\mathcal{R}(\theta)$ to $\mathcal{R}^*(\theta)$ is achieved by redistributing the shares from the upper class towards the middle class without modifying either the mean in $\theta \in [\theta_p, \infty)$, Θ, or the number of poor, $\mathcal{R}(\theta_p)$. The effect of the poor entering the market is studied later.

Two possible cases arise depending on the initial θ_q and, in turn, the initial degree of inequality. Recall that $\Theta = \int_{\theta_p}^{\theta_q} \theta \, \mathrm{d}\mathcal{R}(\theta) + \int_{\theta_q}^{\infty} \theta \, \mathrm{d}\mathcal{R}(\theta)$, so if high values of $\theta > \theta_q$ are found with positive probability, the more unequal the society and the higher the odds of having $\theta_q < \Theta$. In the opposite case, the values of θ are more concentrated around θ_q. As a result, the shareholders are more homogenous and $\theta_q > \Theta$ is a more likely outcome.

If $\theta_q < \Theta$, the initial situation implies high inequality within the upper class and between it and the middle class. As shown in Figure 1(a) after redistribution, $\mathcal{R}^*(\theta_q) > \mathcal{R}(\theta_q)$ (i.e. $1 - N_q^* > 1 - N_q$) so the fixed break-even demand of the marginal firm N_q falls, increasing the number of firms operating formally. As expected, redistribution raises the share of profits held by the middle class, $\Theta_q^* > \Theta_q$, and, in doing so, raises the profit multiplier in (9). Notably, initial high inequality implies that after redistribution, a sizeable upper class still exists, so the entry costs of formal firms are still covered while sales from informal firms decrease. As Figure 1(b) shows, the F-curve shifts up from F_A to F_B. Given the higher

profits and lower N_q in the new equilibrium, the size of the informal economy declines[20].

On the other hand, redistribution under $\theta_q > \Theta$ renders the opposite outcome: it raises N_q and lowers Θ_q, so the F-curve falls from F_B to F_A in Figure 1(b). Initially, the upper class was close to the middle class, so the natural consequence of a drop in the former makes firms use part of their revenues from the middle class to cover their costs after redistribution. As a result, profits decrease, leading to a larger informal sector.

3.2 Providing the public good

Consider an increase in the tax rate, $\mathrm{d}\tau > 0$. This shock, depending on its effects on the tax base q, can either increase or decrease tax revenues and fiscal expenditure (since $g = \tau q$). The latter response is crucial for the following analysis as it defines the direction in which ρ moves and, subsequently, the movements of the M- and F-curves.

Suppose first that $\mathrm{d}\tau > 0$ and $\mathrm{d}g < 0$, which implies that $(\mathrm{d}\tau/\tau) < -(\mathrm{d}q/q)$ and, importantly, that $\mathrm{d}\rho < 0$. This situation is shown in Figure 2(a). The M-curve shifts to the right (from M_0 to M_1) as a consequence of the higher entry costs that are reinforced by a decline in the benefits of formality, ρ. As the reduction in ρ shrinks the profit multiplier in (9), the F-curve contracts from F_0 to F_1. The new equilibrium, with higher N_q and lower π, is characterised by a smaller number of formal sectors ($\mathrm{d}q < 0$) and lower sales both in the formal sector ($\mathrm{d}S^F < 0$) and in the whole economy ($\mathrm{d}S < 0$). The effect on the relative size of the informal sector is, however, ambiguous. The labour demand of formal firms decreases in response to lower sales but increases as the unit labour requirement $\alpha(1-\rho)$ becomes higher. If the first effect dominates, then the informal share of the economy increases ($\mathrm{d}L^I = \mathrm{d}S^I > 0$).

For this case to be feasible, q must be very responsive to the tax change. This is likely to happen when N_q is initially small, which implies a high value for θ_q. As seen in the previous subsection, this setup is consistent with a large middle class in conjunction with a reduced upper class, i.e. low initial inequality. Hence, the drop in sales following the reduction in the income (as $\mathrm{d}\pi < 0$) appears to dominate the needs for more labour in the formal sector, implying a sharp contraction of the F-curve and a larger informal sector[21].

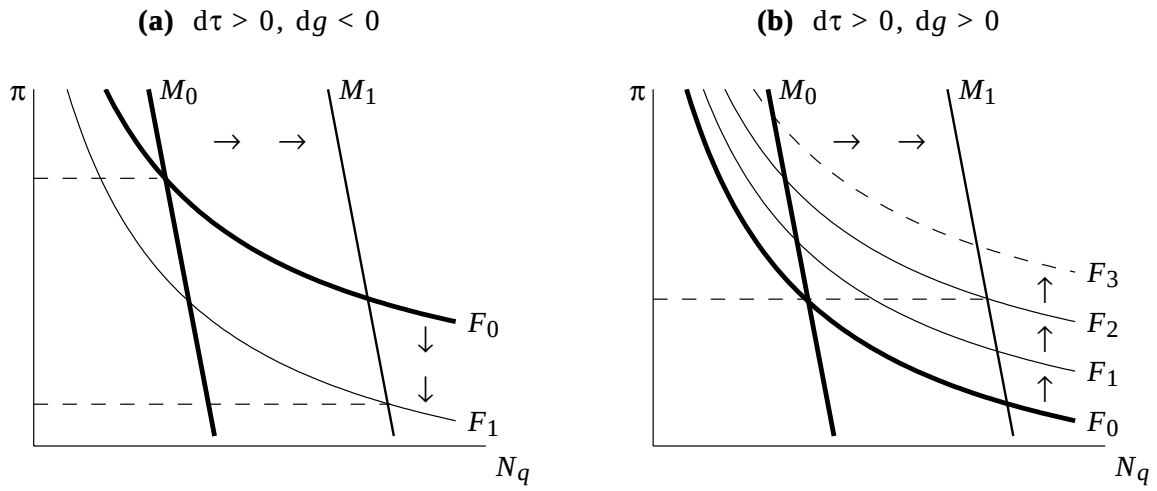
Consider now that $\mathrm{d}\tau > 0$ and $\mathrm{d}g > 0$, which implies that $(\mathrm{d}\tau/\tau) > -(\mathrm{d}q/q)$ and $\mathrm{d}\rho > 0$. As shown in Figure 2(b), the M-curve shifts to the right (from M_0 to M_1) again[22] but the

[20] The M-curve may shift to the left as a result of a higher ρ due to the effects of a larger formal sector on tax revenues. The direction of this change reinforces the expansion of the F-curve and the conclusion is the same.

[21] This is the static comparative result of models with no distributional features as Rauch (1991).

[22] Note that the percentage increase in τ is greater than that of ρ. The M-curve shifts to the right as long as $(\mathrm{d}\tau/\tau) > -\varepsilon_g(1-\varepsilon_g)(\mathrm{d}q/q)$, which is satisfied if $(\mathrm{d}\tau/\tau) > -(\mathrm{d}q/q)$ holds.

Figure 2. Tax increase to finance the public service



F-curve expands, delivering three possible outcomes. If the shift is from F_0 to F_1, in the new equilibrium the drop in profits translates into a smaller number of formal industries, lower overall sales, and lower formal sector sales ($\mathrm{d}q$, $\mathrm{d}S$ and $\mathrm{d}S^F$ are all negative). However, the drop in S^F together with the higher benefit from being formal (lower input requirement, $\alpha(1-\rho)$ since $\mathrm{d}\rho > 0$) leads unambiguously to a lower labour demand from formal firms and a larger operating informal sector[23].

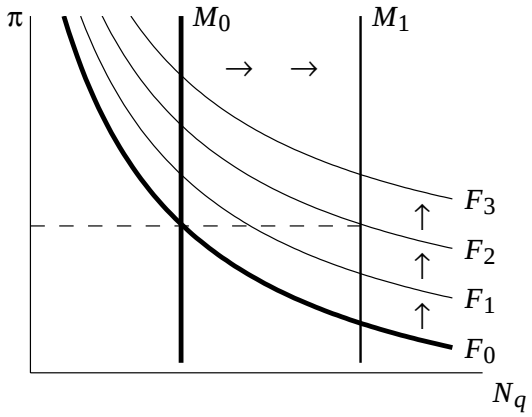
The equilibrium resulting from the shift from F_0 to F_2 (where $\mathrm{d}\pi = 0$) is qualitatively similar to the one just described, with the important difference that $\mathrm{d}S = 0$, so that the change in the size of the informal sector is, albeit positive, smaller. Notice that, *ceteris paribus*, the value of Θ_q implied in the curve F_2 is higher than the one associated with F_1. Hence, a larger middle class or lower inequality, by increasing the multiplier, makes the (expansionary) effects of a larger provision of public good (more fiscal resources, in general) more powerful.

This fact opens the possibility of ending up in the intersection of M_1 with an F-curve above F_2 (for instance, F_3). This equilibrium will have different properties from the previous ones as the tax increase raises both N_q and π, and, notably, it may have a smaller informal sector as an outcome. For a given Θ_q, consistent mainly with $(\mathrm{d}\tau/\tau) > -(\mathrm{d}q/q)$, this case may arise if ρ is large enough, i.e. if the public good or service is fairly productive[24].

[23] Interestingly, the same result arises in Loayza (1996), where the public good is subject to congestion and there is free-riding from the informal firms in its use.

[24] The positive relationship between τ and the size of the informal sector is in line with the empirical finding in Friedman *et al.* (2000) discussed in footnote 3. It may prove useful to see whether the particular sample used in their analysis is driving this result, notably the inclusion of Scandinavian countries or others where the government share in spending is considerably large.

Figure 3. Tax increase to finance the transfers program



3.3 Transfers

In the model, an exogenous increase in the number of shareholders N_p expands the F-curve without perturbing the M-curve. The outcome is a lower level N_q, higher profits and a smaller informal sector. The reason is that as the number of market participants rises, the demand of lower-indexed goods expands thus rendering higher profits to these sectors. The multiplier effect does the rest of the work. With this in mind, I now analyse the workings of a redistributive use of tax revenues: a transfers program. The aim is to link the number of participants to tax revenues in order to study demand effects from fiscal revenues, as opposed to the supply effects considered so far. For this reason, it is assumed that the government allocates a fixed amount g for the provision of the public good, so ρ becomes constant (which is equivalent to set $\varepsilon_g = 0$).

As discussed before, there are $\mathcal{R}(\theta_p)$ people whose incomes are lower than the price of the traded good. The transfers program bridges the income gap $w(\alpha - 1)$ for n people as far as the amount of tax collection allows. The government's new budget constraint, in place of (4), is

$$g + (\alpha - 1)n = \tau q \tag{10}$$

and the number of participants equals the number of shareholders plus the transfers receivers, $N_p = 1 - \mathcal{R}(\theta_p) + n$[25].

In this new setup, the M-curve becomes vertical, so the tax rate τ determines the equilibrium level of N_q by itself. The F-curve remains negatively-sloped. Figure 3 displays the effects on a tax increase used to finance the transfers program. The characteristics of

[25] Notice that the higher purchasing power of these n households does not affect the multiplier as Θ_q remains unchanged.

the equilibria given by the intersection of M_1 and F_1 or F_2 are similar to the analysis of Figure 2(b): the informal sector enlarges following the shock.

The rationale of the transfers program is to increase the demand of lower-indexed goods. Hence, a sufficiently large expansion of the F-curve to reach F_3 requires the introduction of a large number of people to the market, which is a likely situation if the initial poor rate is moderately high. This implies that the program induces a significant increase in the size of the middle class. It is possible, once again, to observe a reduction in the informal economy as a consequence of a higher tax rate as it increases tax revenues.

3.4 Empirical hypotheses

The model provides two empirically testable predictions regarding the effects of income distribution on the size of the informal sector in an economy. Firstly, *income distribution affects informality directly*, as it allocates resources between those who purchase only formal goods and those who also consume informal goods. Specifically, high inequality leads to a large informal sector. By the same token, if inequality is high or, alternatively, the amount of wealth concentrated in the upper class is disproportionately large, redistribution towards the middle class reduces informality. On the contrary, if inequality is relatively low, redistribution may weaken the upper class and reduce the base spending that pushes firms to formality.

Secondly, *income distribution also affects the way informality responds to fiscal changes* that affect either the demand or the supply side of the economy. In particular, the more unequal a society, the less powerful the fiscal policy. As the middle class expands, the marginal effects of a further fiscal intervention against informality are reinforced. This analysis suggests therefore that redistributive policies promoted by the government may trigger a virtue circle in reducing the size of the informal sector.

I now ascertain whether the data support these hypotheses.

4 Empirical evidence from urban Mexico

As mentioned earlier, Chong and Gradstein (2004) provide empirical support to the fact that inequality measures, such as Gini or Theil coefficients, affect positively the size of the informal sector within countries. In this section I perform a parallel analysis using data from Mexican cities during the '90s and the early 2000s.

4.1 A glimpse of Mexico

The Mexican case nicely suits the purpose of testing the predictions of the model for various reasons. In Mexico income distribution is mostly unequal, with a Gini coefficient around 0.5, and the informal sector represents about 30 percent of GDP. Nonetheless, as with among different countries, cross-sectional variation across Mexican states and cities can be considerable. As stressed in Chiquiar (2005), the development experience of the 32 Mexican states has been dissimilar, with the historically wealthier northern states enjoying higher growth rates of per capita income than the relatively poorer southern states[26].

Time variation can also be found as some important macroeconomic events in the 90s had different effects on income dynamics in the different regions. Lopez-Acevedo and Salinas (2000) show that the financial crisis after the peso devaluation in December 1994 reduced income inequality by depressing the labour earnings of the highly-skilled workers (the Gini coefficient fell from 0.534 in 1994 to 0.519 in 1996); once the economy recovered from the crisis, inequality increased. A second major event was the enactment of the NAFTA also in 1994. Cortez (2001) and Chiquiar (2005) suggest that the trade agreement led to a rapid growth in export-orientated manufactures, widening the wage gap between skilled and unskilled work. Consequently, the states with high industrial participation in local GDP underwent an increasing wage inequality as opposed to those producing mainly non-tradables as services.

Another reason that makes the Mexican case appealing is that even though the degree of law enforcement and some legal practices may vary across states, personal and corporate income taxes and social security contributions faced by the private sector are roughly the same across them[27]. This is an important advantage over cross-country studies where there is a stronger need to control for these factors, most of the time with very imperfect measures (i.e. indices) of fiscal burden or regulation depth. For a state or city, those factors can be regarded as fixed, whereas major tax reforms or changes in the rules of the game (for instance, as a result of the 1994 events) can be largely treated as time effects.

A crucial point to reinforce the above ideas and correctly interpret the empirical results below is to understand the way the various levels of government relate in fiscal matters. Tax collection and administration are centralised by the Federal government and the financial dependence of states accounts on federal transfers (*Fondos de Aportaciones*) is considerably

[26] As a matter of fact, the latter are ethnolinguistically more heterogenous than the northern states, with a significant share of indigenous population. La Porta *et al.* (1999) find that, among countries, this fractionalisation affects the performance of government and productivity.

[27] Of course, there are some state level taxes but factor mobility and its limited share in the state revenues (see next paragraph) make them unimportant for the analysis. There are also differentiated taxes aimed to encourage sectoral or regional investment or trade. For instance, the VAT rate is 15% in all Mexico except in border areas where it is 10%. In the empirical work the accounting for the sectoral structure or the geographical location of states controls for these sources of heterogeneity in the tax system.

large. By the late 90s for instance, more than 80% of the states' expenditures was financed by federal resources and just about 6% by their own tax revenues. Clearly, fluctuations in the fiscal accounts at the state level are mainly due to changes in the federal budget, i.e. aggregate shocks. Thus, the correlation between fiscal revenues and expenditures at the state level is not as high as postulated in section 3. So, from now on, it would be useful to consider changes in the fiscal stance not as changes in τ but in g directly.

4.2 Data

An unbalanced panel using city and state variables was built for the years 1992, 1994, 1996, 1998, 2000 and 2002. When data are only available at the state level, the same figure is assigned to cities within the same state . The main source of information is the Mexican National Statistical Office, INEGI, with the exception of data on inflation which come from the Bank of Mexico[28] and information on fiscal accounts that comes from the Centre for Public Finance Studies of the Chamber of Deputies[29].

Employment-related variables come from the National Survey of Urban Employment, ENEU, which is the basis for calculating official labour statistics. In 1992, the ENEU provided data from 32 major metropolitan areas; by 2002 it covered 48 cities[30]. Income distribution and inequality measures come from the National Survey of Household Income and Expenditure, ENIGH, which provides more detailed and complete information on income than the ENEU, which only supplies information on reported wages[31].

Two different operational criteria for identifying people working in the informal sector are common in applied work[32]. The first, pioneered by the International Labour Organization (labelled *Size*), counts as informal those who are employed in firms with five or fewer employees. Clearly, this count would overstate the actual size of the informal sector. The fact that a small firm tend to be informal does not mean that every small firm is informal, whereas self-employed people can easily comply with the prevailing taxation and regulation. The second criterion counts as formal those who receive social security coverage or pay social security taxes as a condition of employment. Then, informality is approximated by the proportion of occupied people who do not receive any social benefit but do *earn a wage*

[28] See *http://www.banxico.org.mx/siteBanxicoINGLES/index.html*.
[29] See *http://www.cefp.org.mx*.
[30] The sample is selected to be geographically and socioeconomically representative. Further details on the evolution of the ENEU can be found at *http://www.inegi.gob.mx*.
[31] A word of caution, though: only the 1998 wave of the ENIGH is representative at the state level (all waves are representative at the national, urban and rural levels). The remaining waves are representative for a varying subset of the 32 states. Therefore, I do not use the expansion factors when computing the inequality measures (results with weighted figures are similar, though). By the same token, the statistics based on ENIGH should be viewed as indicative.
[32] See for instance Marcouiller *et al.* (1997) and Schneider and Enste (2002, ch. 5).

Table 1. Informality and income distribution in Mexican states

	Informality, Social security			*Informality, Size*		
	1992/94	1996/98	2000/02	1992/94	1996/98	2000/02
Poorest 5	29.7 (6.3)	34.4 (7.3)	31.5 (6.2)	48.8 (5.9)	51.2 (5.7)	49.9 (6.5)
Richest 5	19.7 (5.9)	21.4 (5.8)	20.0 (5.8)	38.9 (5.2)	36.6 (6.0)	35.3 (6.1)
All	23.3 (6.7)	27.6 (7.5)	24.0 (7.0)	43.7 (6.1)	44.2 (7.1)	42.0 (7.6)
	Income dist., 100·Gini coefficient			*Income dist., Quintiles 3 plus 4*		
	1992/94	1996/98	2000/02	1992/94	1996/98	2000/02
Poorest 5	53.6 (5.9)	51.8 (4.4)	55.9 (6.1)	31.4 (2.3)	33.3 (2.4)	32.5 (4.3)
Richest 5	48.8 (3.3)	49.7 (5.2)	49.4 (4.5)	33.2 (2.8)	32.6 (3.8)	32.9 (2.2)
All	49.6 (5.6)	51.6 (6.6)	50.1 (6.0)	33.1 (3.1)	31.8 (4.2)	33.2 (3.6)

The table displays sample averages and standard deviations in parentheses. The states were ranked according to their GDP per capita in 1992. The measures of the informal sector come from ENEU while those related to the income distribution come from ENIGH (figures from ENIGH are not weighted, see footnote 31).

(labelled *Social Security*). Note that this follows the usual practice of excluding domestic servants and people with other types of jobs that may be rewarded partially in kind. For this reason, this measure may understate the size of the informal sector.

Table 1 displays descriptive statistics of the above measures of informality. Unsurprisingly, the figures of the Social Security criterion are smaller than those corresponding to the Size criterion. The former are closer to independent estimates of the size of the informal sector, around 30 percent[33]. At the aggregate level, both measures display similar dynamics, peaking by the mid-'90s and falling towards the early-2000s. This pattern is viewed only among the poorest states when using the Size criterion. This casual inspection of the data suggests that the Size criterion may not only overstate the size of the informal sector but may also capture the dynamics of other labour market issues rather than working in the informal sector. For robustness' sake, I consider both measures in the regressions below, but the Social Security criterion, the preferred measure, will receive most of the attention.

The table also reports the Gini coefficients and the share of total income in the hands of the third and fourth quintiles (of the income distribution) as an approximation of the size of the middle class in each state[34]. Clearly, most of the changes have been observed in the poorest states, precisely those with higher and more volatile measures of informality. At first glance, the cross-sectional differences appear to be supportive of the predictions of the model. The conclusion, nonetheless, is not as neat when tracking the time variation.

[33] See Schneider and Enste (2000, 2002) and Schneider (2005).

[34] Similar analysis with alternative measures of inequalities (Theil index and other entrophy measures) and various measures of the size of the middle class (different quintiles and combinations of them) were performed (but not reported). The results barely differ from the ones reported in Table 2.

4.3 Methodological issues and results

Next I perform a set of linear regressions to explain the size of the informal sector using information at the city and state levels. At the city level, I control the estimations for price level inflation and for the fraction of employed working in the US. This last control accounts for the scope of the labour mobility from Mexico to the US which has been intensified as a result of the NAFTA and, of course, for geographical differences between the border cities and the rest.

It is evident from Table 1 that the level of income influences informality. Hence, the regressions include the state GDP per capita which may be thought of as the empirical counterpart of N_p, the number of shareholders in the theoretical model that is negatively correlated with poverty. Nevertheless, for some states the GDP per capita is far from being a good measure of income or development. This is the case of the states where a very large but also very volatile share of GDP is generated from the exploitation of oil reserves by the government-owned oil company (PEMEX)[35]. To address this issue I include explicitly this share in the regressions. Additionally the share of services (excluding financial services) in the state GDP is considered, as it is normally the less capital-intensive sector and the most likely to host informal firms.

In virtue of the fiscal dependence of states explained in section 4.1, I include the ratio of expenditure to own state revenues as an indicator of both the provision and maintenance of public goods and services as well as direct transfers for social programs. Furthermore, this variable interacts with the Gini coefficient or the size of the middle class to test whether income distribution affects the effects of fiscal resources availability on the size of the informal sector.

I first estimate the regression model with fixed effects OLS. Although this approach correctly controls for unobservable fixed factors or year effects, it may render biased estimates as the right-hand-side variables are probably endogenous. Hence, an instrumental variables procedure is required, and I run fixed effects IV using the following instruments: the age dependency ratio (dependents to working-age population); the fraction of indigenous population; the number of homicides and people with pending judicial processes per 10,000 habitants; weather indicators; and the first lag of the explanatory variables[36].

Unfortunately, the available instruments are less than perfect and there are some chances of having a weak instrument problem. Taking advantage of the dynamic nature of the data, the third estimation approach is to add a full set of instruments consisting of suitable lags of the levels of the explanatory and dependent variables for each cross-section involved in the panel and then to perform a GMM estimation. This is the well-known dynamic panel

[35] For instance, the states of Campeche and Tabasco.
[36] The age dependency ratio is available at a city level from ENEU.

data estimator developed in Arellano and Bond (1991)[37].

Table 2 presents the estimations results. To save space, only the GMM estimations in which informality is measured according to the Size criterion are reported. For each estimation method, two regressions were run. The first (odd columns in the table) uses the Gini coefficient to summarise the income distribution, while the second (even columns) uses the proxy for the size of the middle class.

It can be observed that a higher GDP per capita leads significantly to a smaller informal sector, whereas a higher share of services in the local GDP increases informality. In addition, an increase in the expenditure to own revenues ratio, i.e. a larger transfer from the federal to the state government, reduces informality. These results are robust to the measure of informality or the estimation method used.

Consider the regressions with the Gini coefficient as the explanatory variable. The coefficient associated with this index is positive, which implies that higher inequality tends to increase the size of the informal sector, as predicted by the theory. Moreover, the coefficient of the interaction term is also positive, which implies that higher inequality *reduces* the effect of more fiscal resources on informality (which is negative). This finding is also supportive of the predictions of the model. Nonetheless, these results are to be handled with care as they are only significant at a 10% level.

When considering the size of the middle class as a regressor, the evidence in favour of the theoretical predictions is strengthened. A higher middle class implies a lower extent of informality. This is true even after controlling for income levels (i.e. GDP per capita) so a redistribution towards a larger middle class reduces the size of the informal sector. The high inequality observed in the Mexican case is likely to be driving this finding. In terms of the discussion in section 3.1, a larger middle class keeping poverty constant implies a reduction in the upper class and a demand expansion in local markets, while the upper class after redistribution is still large enough to cover the fixed costs of local firms.

Additionally, in this case the coefficient of the interaction term is negative, suggesting that a more concentrated income distribution towards the middle class *reinforces* the negative effect of fiscal stimuli on informality. Consistent with the model, this result can be read as an increase in the profit multiplier (which pushes the F-curve up). These results are significant at least at a 5% level.

[37] The results were similar after applying the extended GMM estimator in Arellano and Bover (1995).

Table 2. Estimation results

	Social security criterion						*Size criterion*	
	FE, OLS		FE, IV		GMM		GMM	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged informal sector size					–0.103*	–0.124*	0.136**	0.145**
					(0.042)	(0.053)	(0.068)	(0.071)
Income distribution (A)								
Gini coefficient	5.102**		7.567**		5.609**		7.086**	
	(2.857)		(3.317)		(2.622)		(0.710)	
Middle class size		–4.560**		–6.478*		–5.193*		–6.466*
		(2.121)		(1.868)		(1.743)		(3.162)
City level variables								
Inflation	–0.196	–0.125	0.669*	0.991*	0.263**	0.247**	0.088	0.056
	(0.179)	(0.177)	(0.272)	(0.243)	(0.141)	(0.115)	(0.133)	(0.131)
Employed in the US (%)	–3.086*	–3.024*	–1.084	–1.761	–0.843**	–0.473	–0.265*	–0.367
	(0.640)	(0.646)	(1.039)	(1.499)	(0.277)	(0.361)	(0.083)	(0.821)
State level variables								
GDP per capita (100·log)	–0.095*	–0.097*	–0.250*	–0.191*	–0.097*	–0.137*	–0.071**	–0.080*
	(0.038)	(0.046)	(0.042)	(0.045)	(0.041)	(0.039)	(0.043)	(0.042)
GDP mining and oil (%)	0.354*	0.354*	0.337**	0.400**	0.437*	0.462*	0.209**	0.198**
	(0.159)	(0.161)	(0.193)	(0.202)	(0.147)	(0.1445)	(0.121)	(0.109)
GDP services (%)	0.384*	0.377*	0.343*	0.366*	0.306*	0.296*	0.742*	0.773*
	(0.081)	(0.087)	(0.111)	(0.117)	(0.087)	(0.085)	(0.069)	(0.702)
Expenditure/Revenues (B)	–0.152*	–0.124**	–0.108**	–0.186*	–0.090**	–0.189*	–0.197**	–0.384*
	(0.064)	(0.059)	(0.044)	(0.066)	(0.042)	(0.058)	(0.088)	(0.161)
Interaction								
$A \cdot B$	0.133**	–0.183**	0.164**	–0.234*	0.224**	–0.297*	0.516**	–0.285*
	(0.061)	(0.091)	(0.078)	(0.085)	(0.114)	(0.094)	(0.310)	(0.099)
Observations	235	235	203	203	152	152	152	152
Adjusted R^2	0.235	0.270	0.293	0.338				
Sargan test (p-value)					0.316	0.662	0.587	0.815
AR(2) test (p-value)					0.752	0.789	0.447	0.454

All regressions include a set of year dummies. The figures in parentheses are standard errors: robust for OLS and IV and two-step corrected for finite sample bias for GMM (Windmeijer, 2005). * [**] denotes significance at a 5% [10%] level. The Sargan test for overidentifying restrictions is asymptotically distributed as χ^2_k under the null of instrument validity, being k the number of overidentifying restrictions. The AR(2) tests for serial autocorrelation of residuals is asymptotically $N(0,1)$ under the null of no serial autocorrelation (Arellano and Bond, 1991).

5 Summary and conclusions

Government regulation, taxation, and the institutional quality in a country have been traditionally regarded as the main causes of informality. Yet some alternative approaches have emerged to explain this widespread phenomenon by relying on the way market behaves and its frictions. This study aims to contribute with this fresh view of the informal sector by addressing the question of whether income inequality affects informality and by focusing exclusively on its effect on market demand.

A simple model with features from Dasgupta and Ray (1986), Murphy *et al.* (1989) and Matsuyama (2000) is developed. On the production side, firms decide simultaneously the quantity of output to produce and the sector in which to operate. Each sector embodies a different technology. Formal firms produce under increasing returns to scale as they can use a productive public services in exchange of a tax payment. Informal firms, on the other hand, neither pay taxes nor access to the public services benefits. It follows that it is easier to comply with the prevailing regulations if the demand firms face for a given fiscal burden (tax rate) is large enough. In other words, the larger the demand that a firm has to meet, the higher the benefits from formality. This fact establishes a link with the demand side on the model which is entirely determined by the income distribution.

With this framework we were able to find that high inequality leads to a large informal sector, and that redistribution towards the middle class decreases the size of the informal sector while it increases the ability of fiscal instruments, either demand or supply-orientated, to reduce informality. To support this, empirical evidence for Mexican cities is provided.

The analysis can be extended in various ways, but I want to emphasise one particular extension. I have refrained from analysing any feedback from informality to income distribution, which might be an important building block to improve our understanding of the workings and effects of the informal sector. This could be achieved by introducing a market for shares or heterogeneity among workers and earnings in the formal and the informal sector in the model. If, as in Rauch (1993), we accept that informal workers are unskilled and have limited access to the stock market, then a causality from informality to poverty and inequality would arise[38]. This channel may set the basis for moving to a dynamic setup where income distribution affects informality and vice versa. This seems to be a promising direction for future research.

[38] The effect of informality on poverty and inequality is not *a priori* clear. One can argue that informality prevents achieving optimal capital-labour mixes and thus is likely to increase poverty and inequality. However, one can also postulate that when the informal sector is large, a person who is not employed in the formal sector can easily enter the informal sector which may be a better situation than becoming unemployed. Hence, informality may buffer the drop in that individual's income and, upon aggregation, mitigate poverty and inequality.

References

Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economics Studies*, 58(2), 277–297.

Arellano, M. and O. Bover (1995), "Another look at the instrumental variable estimation of error-components models", *Journal of Econometrics*, 68(1), 29–52.

Antunes, A. R. and T. V. Cavalcanti (2004), "Accounting for the hidden economy: Barriers to legality and legal failures", Banco de Portugal Working Paper 20-04.

Asea, P. K. (1996), "The informal sector: Baby or bath water? A comment", *Carneige-Rochester Conference Series on Public Policy*, 45, 163–171.

Barro, R. J. and X. Sala-i-Martin (1992), "Public finance in models of economic growth", *Review of Economic Studies*, 59(4), 645–661.

Benassi, C., A. Chirco and M. Scrimitore (2002), "Income concetration and market demand", *Oxford Economic Papers*, 54(4), 584–596.

Chiquiar, D. (2005), "Why Mexico's regional income convergence broke down", *Journal of Development Economics*, 77(1), 257–275.

Chong, A. and M. Gradstein (2004), "Inequality, institutions and informality", InterAmerican Development Bank Working Paper 516.

Cortez, W. W. (2001), "What is behind increasing wage inequality in Mexico?", *World Development*, 29(11), 1905–1922.

Dasgupta, P. and D. Ray (1986), "Inequality as a determinant of malnutrition and unemployment: Theory", *Economic Journal*, 96(384), 1011–1034.

De Soto, H. (1989), *The other path: The invisible revolution in the Third World*, London: IB Tauris.

Friedman, E., S. Johnson, D. Kaufmann and P. Zoido-Lobaton (2000), "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries", *Journal of Public Economics*, 76(3), pp. 459–493.

Johnson, S., D. Kaufmann and A. Shleifer (1997), "The unofficial economy in transition", *Brooking Papers on Economic Activity*, 2, pp. 159–239.

Johnson, S., D. Kaufmann and P. Zoido-Lobaton (1998), "Regulatory discretion and the unofficial economy", *American Economic Review: Papers and Proceedings*, 88(2), pp. 387–392.

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

Levenson, A. R. and W. R. Maloney (1998), "The informal sector, firm dynamics and institutional participation", World Bank Policy Reserach Working Paper 1988.

Loayza, N. V. (1996), "The economics of the informal sector: A simple model and some empirical evidence from Latin America", *Carneige-Rochester Conference Series on Public Policy*, 45, 129–162.

Lopez-Acevedo, G. and A. Salinas (2000), "How Mexico's financial crisis affected income distribution", World Bank Policy Research Working Paper No. 2406.

Marcouiller, D., V. Ruiz de Castilla and C. Woodruff (1997), "Formal measures of the informal sector wage gap in Mexico, El Salvador and Peru", *Economic Development and Cultural Change*, 45(2), 367–392.

Matsuyama, K. (2000), "A Ricardian model with a continuum of goods under nonhomothetic preferences: Demand complementarities, income distribution and North-South trade", *Journal of Political Economy*, 108(6), 1093–1120.

Murphy, K. M., A. Shleifer and R. Vishny (1989), "Income distribution, market size and industralization", *Quarterly Journal of Economics*, 104(3), 537–564.

Rauch, J. E. (1991), "Modeling the informal sector formally", *Journal of Development Economics*, 35(1), 33–47.

Rauch, J. E. (1993), "Economic development, urban underemployment and income inequality", *Canadian Journal of Economics*, 26(4), 901–918.

Robinson, J. (1969), *The economics of imperfect competition*, 2nd edition, London: Macmillan.

Rosser, J. B., M. V. Rosser and E. Ahmed (2000), "Income inequality and the informal economy in transition economies", *Journal of Comparative Economics*, 28(1), 156–171.

Rosser, J. B., M. V. Rosser and E. Ahmed (2003), "Multiple unofficial economy equilibria and income distribution dynamics in systemic transition", *Journal of Post Keynesian Economics*, 25(3), 425–447.

Schneider, F. and D. H. Enste (2000), "Shadow economies: Size, causes and consequences", *Journal of Economic Literature*, 38(1), 77–114.

Schneider, F. and D. H. Enste (2002), *The shadow economy: An international survey*, Cambridge, UK: Cambridge Univeristy Press.

Schneider, F. (2005), "Shadow economies around the world: what do we really know?", *European Journal of Political Economy*, in press.

Windmeijer, F. (2005), "A finite sample correction for the variance of linear efficient two-step GMM estimators", *Journal of Econometrics*, 126(1), 25–51.

A Differential system

Totally differentiating the M- and F- curves leads to

$$\left(\frac{1}{N_q}\right) \mathrm{d}N_q = \left(\frac{1}{\tau}\right) \mathrm{d}\tau - \left(\frac{1}{\alpha}\right) \mathrm{d}\alpha - \left(\frac{1}{\rho}\right) \mathrm{d}\rho \tag{A1}$$

and

$$\left(\frac{1}{\pi}\right) \mathrm{d}\pi = \left(\frac{1}{\rho}\right) \mathrm{d}\rho + \left(\frac{1}{1-\rho\Theta_q}\right) (\rho\, \mathrm{d}\Theta_q + \Theta_q\, \mathrm{d}\rho) + \left(\frac{1}{N_p - N_q}\right) (\mathrm{d}N_p - \mathrm{d}N_q) \tag{A2}$$

From (3),

$$\mathrm{d}\theta_q = -\left(\frac{1}{\mathcal{R}'(\theta_q)}\right) \mathrm{d}N_q \tag{A3}$$

From (6),

$$\mathrm{d}q = -\left(\frac{q}{\alpha}\right) \mathrm{d}\alpha + \left(\frac{\theta_q}{\alpha}\right) \mathrm{d}\pi + \left(\frac{\pi}{\alpha}\right) \mathrm{d}\theta_q \tag{A4}$$

By definition, $\Theta_q \equiv \int_{\theta_p}^{\theta_q} \theta\, \mathrm{d}\mathcal{R}(\theta)$ so

$$\mathrm{d}\Theta_q = \theta_p\, \mathrm{d}N_p - \theta_q\, \mathrm{d}N_q \tag{A5}$$

Using the fact that in the model without transfers $g = \tau q$ and $\rho = \rho(g)$, then

$$\left(\frac{1}{\rho}\right) \mathrm{d}\rho = \left(\frac{\varepsilon_g}{g}\right) \mathrm{d}g = \varepsilon_g \left[\left(\frac{1}{\tau}\right) \mathrm{d}\tau + \left(\frac{1}{q}\right) \mathrm{d}q\right] \tag{A6}$$

On the other side, in the model with transfers, see (10),

$$\mathrm{d}N_p = -\left(\frac{n}{\alpha - 1}\right) \mathrm{d}\alpha + \left(\frac{q}{\alpha - 1}\right) \mathrm{d}\tau + \left(\frac{\tau}{\alpha - 1}\right) \mathrm{d}q \tag{A7}$$

Formal sales are given by (7), hence

$$\mathrm{d}S^F = \alpha\, \mathrm{d}N_p + (\Theta_q + \theta_q N_q)\, \mathrm{d}\pi - \left(\frac{\pi N_q}{\mathcal{R}'(\theta_q)}\right) \mathrm{d}N_q \tag{A8}$$

whereas employment in the formal sector is $L^F = (1-\rho)S^F$,

$$\mathrm{d}L^F = (1-\rho)\, \mathrm{d}S^F - S^F\, \mathrm{d}\rho \tag{A9}$$

Since the number of workers is fixed, the change in labour demand in the informal sector (which equals the change in sales in this sector) is

$$\mathrm{d}L^I = \mathrm{d}S^I = -\,\mathrm{d}L^F \tag{A10}$$

The differential version of the overall sales expression (8), where $\Theta \equiv \int_{\theta_p}^{\infty} \theta \, \mathrm{d}\mathcal{R}(\theta)$, is

$$\mathrm{d}S = \alpha \, \mathrm{d}N_p + \Theta \, \mathrm{d}\pi \tag{A11}$$

Hence, the differentials of M- and F-curves in the model with no transfers are

$$\left[\frac{1}{N_q} - \pi\left(\frac{\varepsilon_g}{\alpha q}\right)\left(\frac{1}{\mathcal{R}'(\theta_q)}\right)\right] \mathrm{d}N_q + \theta_q\left(\frac{\varepsilon_g}{\alpha q}\right) \mathrm{d}\pi = (1-\varepsilon_g)\left[\left(\frac{1}{\tau}\right) \mathrm{d}\tau - \left(\frac{1}{\alpha}\right) \mathrm{d}\alpha\right] \tag{A12}$$

and

$$\begin{aligned}\left[\left(\frac{\alpha q}{N_p - N_q}\right) + \left(\frac{1}{1-\rho\Theta_q}\right)\left(\frac{\pi}{\alpha q}\right)\left(\frac{\varepsilon_g}{\mathcal{R}'(\theta_q)}\right)\right] \mathrm{d}N_q + \left[\frac{1}{\pi} - \left(\frac{\varepsilon_g}{1-\rho\Theta_q}\right)\left(\frac{\theta_q}{\alpha q}\right)\right] \mathrm{d}\pi \\ = \left(\frac{\varepsilon_g}{1-\rho\Theta_q}\right)\left[\left(\frac{1}{\tau}\right) \mathrm{d}\tau - \left(\frac{1}{\alpha}\right) \mathrm{d}\alpha\right] + \left(\frac{\alpha}{N_p - N_q}\right) \mathrm{d}N_p\end{aligned} \tag{A13}$$

The slope of the M-curve is

$$\left[\frac{\mathrm{d}\pi}{\mathrm{d}N_q}\right]_M = \frac{1}{\theta_q}\left[\frac{\pi}{\mathcal{R}'(\theta_q)} - \frac{\alpha q}{\varepsilon_g N_q}\right] \leq 0 \tag{A14}$$

which is negative as long as

$$\varepsilon_g \leq \frac{\mathcal{R}'(\theta_q)}{1-\mathcal{R}(\theta_q)}\left(\frac{\alpha q}{\pi}\right) = \frac{\mathcal{R}'(\theta_q)}{1-\mathcal{R}(\theta_q)}\left[\frac{1}{\pi} + \theta_q\right] \tag{A15}$$

Note that the above condition implies a small value for ε_g and, in fact, a steep M-curve. On the other hand, the slope of the F-curve is negative if

$$\varepsilon_g \leq \left(\frac{\alpha q}{\theta_q \pi}\right)(1-\rho\Theta_q) = \left(\frac{\alpha q}{\alpha q - 1}\right)(1-\rho\Theta_q) \tag{A16}$$

which holds if (A15) does.

In the model with transfers, with $\varepsilon_g = 0$ and n endogenous, both curves become

$$\left(\frac{1}{N_q}\right) \mathrm{d}N_q = \left(\frac{1}{\tau}\right) \mathrm{d}\tau - \left(\frac{1}{\alpha}\right) \mathrm{d}\alpha \tag{A17}$$

and

$$\begin{aligned}\left[\alpha q + \left(\frac{\tau}{\alpha - 1}\right)\left(\frac{\pi}{\mathcal{R}'(\theta_q)}\right)\right] \mathrm{d}N_q + \left[\frac{N_p - N_q}{\pi} - \left(\frac{\tau\theta_q}{\alpha - 1}\right)\right] \mathrm{d}\pi \\ = \alpha\left[\left(\frac{q}{\alpha - 1}\right) \mathrm{d}\tau - \left(\frac{1}{\alpha - 1}\right)\left(n + \frac{\tau q}{\alpha}\right) \mathrm{d}\alpha\right]\end{aligned} \tag{A18}$$

so the M-curve is vertical and the F-curve still has negative slope.