Informal firms, migration to formality and government policy

Anders Fredriksson*

June 17, 2007

Abstract

This paper discusses informal firm investment, growth and possible "migration" to formality when there are large entry costs to formality, typical of many developing countries. The questions are addressed in a dynamic model of capital accumulation. Compared to previous studies, the dynamic framework enriches the modelling of the firm decision to become formal or not. High-end infomality may co-exist with a low-end "informality trap".

The transition to formality over time is not obvious. Informal firms may choose to stay small because, in an environment characterized by risk of detection and penalties, it may prove too costly to grow in anticipation of formalization.

The paper also discuss policy vis-a-vis informal firms. Two different views on how policy and policy objectives in the informal economy can be viewed are proposed: the benevolent government and the "malevolent bureaucrat".

Not surprisingly, the "malevolent bureaucrat" may be much detrimental to firm investment and formalization possibilities. A revenue maximizing bribe taker may lead to more firms caught in an "informality trap", compared to the interpretation where the benevolent government has a "carrot and stick"-tool to affect small firm migration incentives.

A characteristic of developing countries is limited enforcement of policy vis-a-vis informal firms. I address this issue by allowing for variation in penalties *and* to what extent the policy maker, either benevolent or malevolent, is actually successful in implementing them.

*IIES, Stockholm University. anders.fredriksson@iies.su.se

1 Introduction

The view that "legal barriers" to doing business may hinder development has been popularized over the last two decades. One such legal barrier is the "regulation of entry" - the costs related to register a new firm such that it complies with all labor-, tax- and other regulations that the authorities require. In Latin America, the average money cost to formalize a business is 944 USD (Doing Business, 2007). With per capita income levels of one tenth to one fifth of the income levels in the US, the cost to formalize a business is substantial. Moreover, income levels in the informal economy are much lower than the official GDP figures, making the money cost to formalize a small informal business very high.

In this paper a dynamic model of capital accumulation is used to study firm investment, growth and possible formalization. Only over time can firms with small capital stocks and limited production reach the size necessary to justify paying the entry cost to formality. Furthermore, the transition to formality over time is not obvious. Informal firms may choose to stay small because, in an environment characterized by risk of detection and penalties, it may prove too costly to grow in anticipation of formalization. Furthermore, even if firms decide to become formal, this may take substantial time because firms need to build the necessary capital stocks.

By using a dynamic model two effects interact in making firms choose formality. The threshold for formalization is related not only to the entry costs but also to how much is to be gained, in our model in terms of productivity, from formalizing the firm. This is the threshold effect. Contrary to static models there is also an accumulation effect, where the lowest-productivity firms can, or cannot, grow to reach the size necessary for formalization.

The second novelty of the paper, in addition to the dynamic setting, is a more detailed modelling of policy vis-a-vis informal firms in developing countries. First, it is not at all obvious which is the policy setting entity relevant for small informal firm production and investment. That is, what government authority and what policy instrument affects the production and investment behavior of small informal firms? We study two different policy setting entities, the benevolent government and the "malevolent" bureaucrat. We use a policy instrument, "penalties", and discuss policy outcomes, i.e. firm investment, growth and formalization behavior, for the two different policy entities proposed. Not surprisingly, results can differ substantially.

Second, there is the policy objective. If a benevolent government indeed has some policy tool to influence the investment and production decisions of small informal firms, what is the policy objective? A complete elimination of informality due to a "competition fairness" or negative externality argument, help firms formalize their business in order to grow or a laissez faire argument have all been proposed in the popular debate.

Embedded in the policy question is the issue of enforcement: to what degree can the policy setting entity actually implement its policy? Informal firms are informal at least in part because the government cannot "reach out" with its policies, something which is more likely to be the case in developing countries. There may be a number of barriers to actual implementation over which the policy maker has no or little control: it may depend on cooperation between the policy-maker and other government agencies, efficient systems to find the informal firms do not exist, an efficient judiciary is lacking etc.

We propose a set-up in which the individual response to penalties is allowed to differ between institutional environments. In an environment where enforcement is difficult there may be much leeway for firms to react to and "hide" from penalties. Alternatively, if the entity setting penalties can enforce the penalties more easily, firms on average have to comply to a larger extent.

A much debated motivation for the existence of the informal economy, also in developing countries, is tax policy. Too high taxes and too little in return for taxes for formal firms is seen as a main determinant of informality. The model presented focuses on the small informal firm that, due to high entry costs, has no access to formality. Taxation and public goods provision act in the background however, by affecting the productivity differential between informality and formality and thereby the incentive to become formal. We discuss this point at the end of our analysis.

Earlier work on informal firm production and formalization use static models, e.g. Rausch (1991), Loayza (1996), Antunes and Cavalcanti (2006) and Prado (2007). There is a literature on poverty traps and nonconvexities that resembles the capital accumulation framework presented here, e.g. Galor and Zeira (1993) and Banerjee and Newman (1994). In terms of policy vis-a-vis informal firms, penalties are normally considered a "tax" on informality, see the static models reference above. The present paper tries to discuss more in detail how policies towards informal firms can be viewed.

The paper is structured as follows. Section 2 discusses entry costs to formality. In section 3 the microfoundations of the model are presented, with an emphasis on the endogenous response of firms to penalties and how to think of government policy, in different environments. We then present the dynamic model of firm investment and growth in section 4, the individual firm formalization decision in section 5 and aggregate behavior in section 6. We then turn to policy in Section 7. Penalty policies under different policy setting entities and policy objectives and "optimal policy" is discussed. The section also briefly discusses taxation and public goods policies. Section 8 concludes.

A note on terminology: Our definition of informal is a firm that has not paid the legal cost of L, in contrast to a formal firm that has gone through this formalization procedure. The decision of the firm to pay the legal cost L and

become formal is called either formalization or "migration" to formality. We use the terms interchangeably .

2 The regulation of entry[1]

The cost of formalizing a business is comprised of both monetary and other costs. It is well documented that these costs can be very high (Djankov et al, 2002). The most up-to date source of information on entry costs is most likely the "Doing Business" project financed by the World Bank. This data set on entry costs originally covered 75 countries (Djankov et al 2002), now 155 countries are included[2]. A summary of the regulation of entry data is presented in table 1, with the number of procedures to register a firm and the official time it takes. The last column measures the official cost of the different registration procedures as a percentage of official GNI. We see that the money cost to start a business is at least 40% of yearly GNI per capita in most of the developing world.

Starting a Business

Region	# Proc	Dur days	Cost %GNI/cap
East Asia & Pacific	8.2	46.3	42.8
Europe & Central Asia	9.4	32.0	14.1
Latin America & Caribbean	10.2	73.3	48.1
Middle East & North Africa	10.3	40.9	74.5
South Asia	7.9	32.5	46.6
Sub-Saharan Africa	11.1	61.8	162.8
OECD: high income	6.2	16.6	5.3
United States	5,0	5,0	0.5

Table 1

Table 2 presents data for the Latin American countries present in the World Bank data, augmented with informal economy income figures from Schneider (2002). 6-17 different bureaucratic procedures with a total cost of 340-1940 USD and consuming 27-152 days have to be taken in order to formalize a firm. The average is 12 procedures, 62 days and 944 USD in official cost. All Latin American countries have higher official cost than the United States and the average cost is 330% of the US cost. Columns 4-7 are the official GNP figures from Schneider (2002), the informal GNP figures from the same source and the ratios between the official cost to the monthly informal GDP and the total

[1] I borrow this term from Djankov et al (2002). Other terms used are "costs of becoming formal" (de Soto) and "starting a business" (World Bank).

[2] Information available on-line at http://www.doingbusiness.org/

cost to the monthly informal GNP respectively[3]. Columns 6 and 7 can thus be interpreted as the number of months an average informal worker would have to work to generate an income equal to the official entry cost and total cost respectively.

If we focus only on the official cost to start a company (column 6) then Argentina and Brazil, the most favorable countries, require almost 3 average informal monthly GNP:s to start a company. The Latin-American average is 16 times informal GNP and Nicaragua has very high costs in terms of informal income. These costs are high or even prohibitive for small informal firms that may not see any clear advantage of formality in a distant future and will make formalizing a firm possible only after a certain capital has been accumulated. Leaving aside the origins of the complex and costly regulatory requirements it is clear that unless formality has something very attractive to offer informality will persist.

Starting a Business and informal GNP

	1	2	3	4	5	6	7
Country	# Proc	Dur days	Off. Cost	Off. GNP/ month	Inf. GNP/ month	Off cost/ GNP/month	Total cost/ GNP/month
Argentina	15	32	540,3	621,7	157,9	3,4	5,9
Bolivia	15	50	1420,2	82,5	55,4	25,7	26,8
Brazil	17	152	341,4	298,3	118,7	2,9	7,1
Chile	9	27	575,4	382,5	75,7	7,6	9,9
Colombia	13	44	452,9	168,3	65,8	6,9	8,6
Costa Rica	11	77	1076,5	317,5	83,2	12,9	16,4
Dom. Rep.	10	73	716,0	177,5	57,0	12,6	15,2
Ecuador	14	65	835,3	100,8	34,7	24,1	26,7
Guatemala	13	30	1250,0	140,0	72,1	17,3	18,5
Honduras	13	44	721,2	71,7	35,6	20,3	21,7
Mexico	8	27	1035,0	422,5	127,2	8,1	9,6
Nicaragua	6	39	1197,1	33,3	15,1	79,5	80,5
Peru	10	72	847,2	173,3	103,8	8,2	9,6
Uruguay	10	43	1928,1	500,0	255,5	7,5	8,7
Venezuela	16	151	1219,5	359,2	120,7	10,1	15,0
Average LA	12	62	943,7	256,6	91,9	16,5	18,7
United States	5	5	285,7				

Table 2

It should be noted at this stage that there may be other ways to start a

[3] The total cost measure, as perceived by an informal agent, is probably a summary measure of the monetary cost + the time cost of actually fulfilling all the requirements + transport costs etc. to visit the different government bodies. The calculation for total cost in column 7 is somewhat ad-hoc and as follows; (the official cost) + (the number of procedures times half the informal average daily GNP) + (an ad-hoc measure of the loss of waiting set to the duration in days divided by 3 times half the daily informal GNP).

The daily GNP is the monthly GNP divided by 20. Each procedure is assumed to require 1 day of work. Each procedure is assumed to have a value of half an average daily informal GNP. The loss due to waiting is set to be a third of the duration time times half the daily GNP.

company and legislation may have been simplified for the smallest firms. It is customary in Brazil to use a "dispatcher" or mediator firm instead of the official process. The dispatcher is contracted to complete all steps necessary for formal registration and full legal status is obtained. One source speaks of an average cost of formalization this way of 640 USD for medium-sized firms (Stone et al., 1996). It is likely that the more complicated the official registration process is the better is the bargaining position of these mediator firms and the higher their price.

As for simplified procedures for the smallest firms some such reforms are under way in for instance Brazil (SEBRAE, 2007). Many Latin American countries have not seen any such reforms though and they are only recently under way in Brazil. It should also be noted at this stage that the tables above indicate how the registration processes work in theory. Reality on the ground may be very different. The most famous example of "how it really works" is probably the de Soto study where it took 289 man-days and cost a total of 32 minimum salaries (including 2 unavoidable bribes) to start up a small firm (de Soto, 1989).

We now turn to our model in which we use one summary measure for the start-up cost of firms, a *legal* cost L. Because many of the the results and the intuition carry through if we include a time cost as well, we work with the simplest case with only a fixed money cost L.

3 Production, policy and enforcement

This section sets up a model to analyze the link between informal firm production, policy vis-a-vis informal firms and the possibilities for enforcement of policies.

Informal means, in the strict meaning of the word, illegal. Firms may be detected and fined for running clandestinely. Below, in the first subsection, we introduce a framework where firms react to the threat of detection and penalties by organizing production in order to minimize such interference. In the second subsection we discuss penalties: what entity sets the "penalty policy" that is relevant for small informal firms and what is the policy objective? In the third subsection we discuss how effective the policy setting entity is in actually implementing its policy.

3.1 Production of the informal firm

Firms own a capital stock k that is used for production and output is linear in the capital stock, Ak. Firms face a risk of being detected and imposed a

fine which we model as a fraction x of the capital stock that, in the case of detection, will have to be paid as a penalty, xk. The "crime" committed by an informal firm of size k is thus considered proportional to its size. We restrict the penalty parameter x to vary between 0 and A, mainly to avoid negative values on production net of penalties.

It is reasonable to assume that actions can be taken such that detection is kept at a low level. Tokman (1992) provides ample evidence that informal firms in Latin America organize part of production so that it is "invisible". The accounts in Tokman contain numerous examples of how small informal firms organize activities to minimize disturbance from authorities, for instance by choosing less visible and less favorable production locations, physically hiding production when authorities visit and in anticipation of such visits, meeting customers one by one due to the lack of a visible sales location and marketing possibilities.

Our modelling of these activities, called "hiding", has three components. First, by spending time on other things than production (i.e. hiding) firms can affect the probability of being detected and fined. Firms optimally choose their time allocation such that production net of penalties is maximized. As a result, changes in penalties result in changes in behavior. Second, we let firms differ in their ability in organizing production such that they are not detected. Firms are thus exogenously different in their ability in avoiding detection. Third, we also allow for the general environment to play a role. That is, we allow the degree to which individual behavior can affect the probability of non-detection to vary.

3.1.1 Maximizing expected production

We assume that both production and the probability of detection (and paying x) is tied to the firm manager´s time. We index firms by j. The firm manager spends a fraction l of her time on affecting the probability of not being detected, defined as $p_j(l)$, and the remaining fraction $(1-l)$ on production. The resulting gross production is $A(1-l)k$. The expected informal firm production net of penalties per unit of capital, i.e. the expected productivity, defined as A_j, then becomes

$$A_j = p_j(l)A(1-l) + (1-p_j(l))(A(1-l)-x) = A(1-l) - x + xp_j(l) \quad (1)$$

Maximizing A_j with respect to time l we get the first order condition

$$x\frac{dp_j}{dl} = A \quad (2)$$

Time l is spent such that the marginal production loss of spending more time on hiding activities, A, equals the production gain from a lower detection probability, $x\,(dp_j/dl)$. Let the probability function take the following form:

$$p(l) = \alpha_j l^e$$

The parameter α_j captures the individual firm´s ability in hiding and we will let α_j be distributed over some interval $0 \leq \alpha_j \leq \alpha_{\max}$. A firm with a low α_j cannot do much to shed itself from detection whereas a high-α_j firm can. The environment parameter e on the other hand describes the general effectiveness of hiding, where e can take values on the unit interval, $0 \leq e \leq 1$. With a small value of e (close to zero) a firm, given its individual hiding capacity, gets a lot out of spending time on hiding. For high values of e (close to 1) however, spending time on hiding is not as effective in avoiding detection.

The environment parameter e can be seen as reflecting how effective collection of penalties is. One can imagine two countries with the same intended penalty policy x but the end result of such a policy, i.e. the collection of penalty payments and resulting firm production, differs a lot. When e is low there is more leeway for individual behavior; firms can more easily "hide" and thereby escape penalties. When e is high, taking individual action to affect the penalty payments will not be as effective, the policy maker can more easily enforce the payments. In what follows we will return to a "low-e" and a "high-e" scenario.

From the first order condition in (2), we get the following individual optimal time allocation l_j^* and resulting productivity parameter A_j^*, where we use $*$ to indicate that l_j^* and A_j^* are the result of a firm optimization decision:

$$\begin{aligned} l_j^* &= \left(\frac{\alpha_j x e}{A}\right)^{1/(1-e)} \\ A_j^* &= A - x + A\left(\frac{\alpha_j x e}{A}\right)^{1/(1-e)}\left(e^{-1} - 1\right) \end{aligned} \tag{3}$$

The optimal firm-specific time allocation l_j^* is increasing in penalties x and firms with higher α_j spend more time "hiding" (because they get more out of it). It is also increasing in e meaning that in environments where it is hard to shed production from penalties firms spend more time doing so. Productivity A_j^* is increasing in α_j, decreasing in penalties x and decreasing in e. That is, penalties affect firm productivity but less so for the firms that have more ability in hiding. Environments in which collection of penalties is less effective (low e) makes for higher productivities because firms get more from their individual hiding abilities.

Expression (3) is the micro tool to analyse effects from penalties on firm productivity. We still need to embed it into a dynamic model of investment, growth and possible formalization however. This is done in section 4 below. At

this point we should note that the time allocation decision of firms is a static problem that can be separated from the dynamic problem of investment, we can thus take A_j^* as given when we set up the dynamic problem.

Before proceeding to the dynamic model however, we turn to the question of what penalties are, what entity sets them and how we can think of the implementation.

3.2 Policy makers and policy goals

We allow for two alternative interpretations of the penalty parameter x. It can be seen as a "carrot and stick" policy tool of a benevolent government that is used in order to affect incentives for investment and formalization of informal firms. A policy objective, maybe due to negative externalities from informality, may be to reduce informality. As a response to a certain penalty rate a firm may choose to invest a lot and try to escape informality as soon as possible. Or it may simply be unable to do so if penalties are set too high. We shall see that, in our dynamic set-up with large entry barriers to formality, there is an interesting interaction between penalties, investment and formalization decisions. Policies that make some firms choose formality make other firms unable to do so.

Another interpretation is that x is a fine to a "malevolent" bureaucracy or police over which the government has little control. The bureaucrat takes bribes or side payments and let firms continue operating clandestinely and the goal of setting x may be to maximize revenue. We can also think of police selling "protection" to small firms against not reporting them. In both these cases we think of the government somewhere in the background with limited control over its bureaucrats/police. Different policies will arise the policy maker can commit to the policy set or simply maximizes current revenue without taking into account the effect on the future capital stock. Not surprisingly, the "malevolent" bureaucracy may be much detrimental to small firm growth.

3.3 Degree of implementation of intended policies

At the core of the discussion of informality and policy is enforcement. A benevolent government may be unable to implement its intended policy because of the general institutional environment. There may be a number of barriers to actual implementation over which the policy maker has no or little control. In terms of penalizing informal firms the implementation of the policy may depend on cooperation between the policy-maker and other government agencies, efficient systems to find the informal firms, an efficient judiciary etc.

In the model presented in this paper, the parameter e represents the institutional environment of a society. In a "low-e" society it is easier for informal firms to escape controls and the firms that are good at doing so are successful in operating informally without paying much penalties. Because of weak governments and limited enforcement, this is likely to be the case in many developing countries. In a "high-e"-society such avoidance of authorities is more difficult.

It should be pointed out that e is not seen as a policy parameter of the model. We rather think of different societies as having different values of e and study how the policy outcome depends on the value e takes.

Before moving to the dynamic problem, we take a simplifying step by considering two different environments instead of the full range $0 \leq e \leq 1$.

3.3.1 Low e

Even small amounts of time dedicated to hiding rather than production affects the probability of detection a lot. Looking at the function that $p(l) = \alpha_j l^e$ we see that for small values of time l the probability will still be close to its (firm specific and exogenously given) maximum value α_j. From (1), with $p_j \approx \alpha_j$ and $l \approx 0$, we should thus expect the resulting firm productivity to be close to $A - x + x\alpha_j$. Indeed, it can be shown that A_j^* approaches this value as $e \to 0$, i.e. we have

$$A_j^* = A - (1 - \alpha_j)\, x \tag{3A}$$

where $0 \leq \alpha_j \leq \alpha_{\max} = 1$.[4] We thus have a case where the firm with no ability in avoiding penalties gets only the "detection productivity" $A - x$ whereas the firm with highest ability in hiding has productivity A.

3.3.2 High e

For higher values of e firms will not get as much out from trying to avoid penalties. This will affect a firm more the better it is at "hiding". In the extreme case of $e \to 1$ all firms get the same productivity $A - x$. Individual responses to penalties have no effect.

In the less extreme case of $e = 1/2$, from (3) we get

$$A_j^* = A - x + \frac{\alpha_j^2 x^2}{4A} \tag{3B}$$

[4] For algebraical simplicity we prefer to work with $e \to 0$ instead of a small but positive value on e, even though in $A_j^* = A - x + x\alpha_j$ there is no real individual response to different penalties because for all penalties l approaches zero.

We get a productivity range from the "detection productivity" $A - x$ to $A - x + x^2/(2A)$ for the range of firms $0 \leq \alpha_j \leq \alpha_{\max} = \sqrt{2}$. If $x = A$ the productivity range is thus from 0 to $A/2$. The following graph shows productivity as a function of the firm-specific α_j for a given high penalty (x close to A), for the low-e and high-e cases[5]:

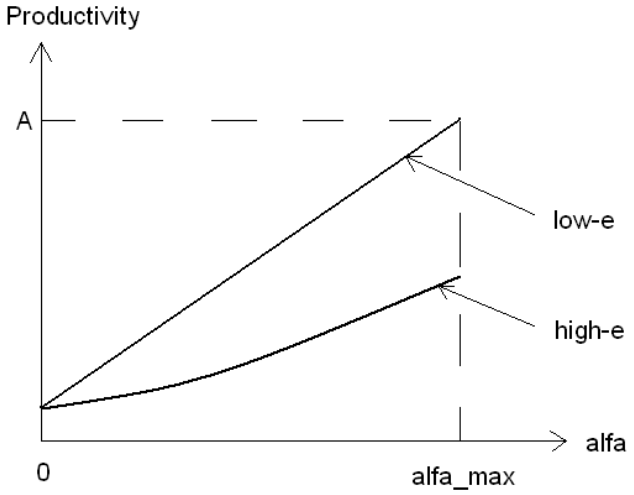

Figure 1

The lower is e the more can firms thus shed themselves from penalty payments, resulting in a higher productivity. Penalty policies will be less effective in achieving a certain goal in societies where enforcement is more difficult[6].

We now turn to the dynamic aspects of the model. For the next two sections we work with the firm specific informal productivity A_j^*, without specifically referring to the micro foundations presented above. The $*$ however indicates that it is the outcome of an individual static time allocation decision.

4 Dynamic problem

The model contains a universe of firms that all start out as informal. The profit-maximizing firms produce output that can either be withdrawn as profits

[5] The reason for $\alpha_{\max}$ differing between the two cases is technical. In a strict sense we start with the assumption of $x \leq A$. Consistent with this assumption we then choose a functional form on $p(l)$ that is as simple as possible. We then let the idiosyncratic firm parameter α_j parameter vary over an interval such that the results are consistent with unit interval restrictions on time spent on hiding and the resulting probability. The maximum α_j that we can then allow depends on the value the e-parameter: $\alpha_{\max} = e^{-e}$. It is a parabola taking the value 1 at $e = 0$ and $e = 1$ and higher values in between, for instance $\sqrt{2}$ at $e = 1/2$.

[6] Implicit in such a reasoning is that the policy maker would like to reduce informality due to for instance a negative externality. We return to this in the policy section.

or invested. Firms have perfect foresight. We use subindex t, for time, on the capital and investment variables. The capital stock k_t of a firm is augmented by investments i_t at a cost of $(z/2)\, i_t^2$ and capital depreciates at the rate δ such that

$$\dot{k}_t = i_t - \delta k_t$$

Firms also need to decide whether or not to become formal, which means paying the formalization payment L. This will be specified in detail below.

4.1 Informality forever

First study the optimal capital accumulation path of a firm with initial capital k_0 that remains informal forever. From its static time allocation decision outlined above, the firm production per unit of capital is A_j^*. The firm chooses its investment path i_t in order to maximize lifetime profits. With per-period profits being production, $A_j^* k_t$, minus cost of investment, $(z/2)\, i_t^2$, the optimization problem is

$$\text{Choose } i_t \text{ to Max } \int_0^\infty \left(A_j^* k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt \qquad \text{s.t.}$$

$$\dot{k}_t = i_t - \delta k_t \qquad k\,(0) = k_0$$

The parameter δ is the depreciation rate and ρ is the rate of time preference. Using the Hamiltonian $H = \left(A_j^* k - \frac{z}{2} i_t^2\right) e^{-\rho t} + \lambda_t \left(i_t - \delta k_t\right)$ we solve for the following investment, capital accumulation and lifetime profit functions:

$$\begin{aligned}
i_{j,t}^{\text{informal}} &= i_j^{\text{informal}} = \frac{A_j^*}{z\,(\delta+\rho)} \\
k_{j,t}^{\text{informal}} &= k_0 e^{-\delta t} + \frac{A_j^*}{z\delta\,(\delta+\rho)}\left(1 - e^{-\delta t}\right) \\
k_{j,\infty}^{\text{informal}} &= \frac{A_j^*}{z\delta\,(\delta+\rho)} \\
\pi_j^{\text{informal}}\,(k_0) &= \frac{A_j^* k_0}{\delta+\rho} + \frac{\left(A_j^*\right)^2}{2z\rho\,(\delta+\rho)^2}
\end{aligned} \tag{4}$$

A few results immediately follow from these expressions. First, investments are constant and the capital grows until per-period investment just replace the depreciated capital each period. Investments, the steady-state capital stock $k_{j,\infty}^{\text{informal}}$ and profits are all increasing in A_j^*. A high discount rate ρ, maybe due to uncertainty about the future, has the opposite effect. So does a high

depreciation rate of capital δ and a high cost of investments z. High values of ρ, δ and z thus all act to use today´s earnings for profits instead of investing in the capital stock to get higher profits tomorrow.

A profit-maximizing ever-informal firm never reaches a capital stock higher than $k_{j,\infty}^{\text{informal}}$. This level may be very low depending on parameters and, intuitively, may be too low for a firm to migrate to formality given that such migration means using part of production for migration costs. We will return to this question in detail.

The initial capital stock affects the total stock of capital and the time it takes to reach a certain level (below $k_{j,\infty}^{\text{informal}}$) but its effect vanishes in the long run. Also, lifetime profits $\pi_j^{\text{informal}}(k_0)$ are affected by the initial capital level.

4.2 Formality

Formality means that a firm can achieve its full production capacity (Ak) because time no longer has to be spent on "hiding"[7]. Formalization requires a payment of a fixed legal cost L however. As argued in section 2 there is evidence that migration costs can be high, especially if compared to the level of capital and profits at which small informal firms act. We will assume that an informal firm can (and does) borrow in order to pay this formalization fee and we assume that it is paid at one single instant in time which makes the firm formal and able to take advantage of formality. We let T be the time at which the firm becomes formal and from this time on per-period interest payments rL have to be paid. The fact that the firm starts paying interest payments rL each period is what matters for when the firm will migrate, prior to migration there is no such cost[8].

Before solving the migration problem we look at a firm once it has become formal. The formal firm has productivity A and needs to decide the optimal investment plan. Production is used for profits, investment and interest payments rL. If having migrated at time T at the migration level of capital, defined as k^{M}, we can treat the problem from T onwards as an infinite horizon problem and discount the calculated profit values by $e^{-\rho T}$ (note that we take k^{M} and T as given at this stage, we do not know them yet and we are to solve for them in what follows). We use the same depreciation rate δ, discount rate ρ and cost of

[7] In the final section of the paper, when discussing policy, we also allow for the possibility that the formal productivity is $A^{\text{formal}} > A$, meaning additional advantages of formality, except for the fact that no time needs to be spent on hiding. For now however, formality means getting productivity A.

[8] The assumption of borrowing L and then paying rL per period is mainly a technical assumption. We have experimented with a three-period model where the firm instead uses part of its capital stock to pay L and the main intuition and comparative statics are similar to the model presented here.

Note that the loan of L is rolled over indefinitely.

investment z as we did for informal firms and we no longer need the subindex j for firm. The problem once formal can thus be written as

$$\text{Choose } i_t \text{ to Max } e^{-\rho T}\int_0^\infty \left(Ak_t - \frac{z}{2}i_t^2 - rL\right)e^{-\rho t}dt \qquad \text{s.t.}$$

$$\dot{k}_t = i_t - \delta k_t \text{ and } k\left(0\right) = k^{\mathrm{M}}$$

The solution is similar to the informal firm problem above;

$$i_t^{\text{formal}} = i^{\text{formal}} = \frac{A}{z\left(\delta+\rho\right)}$$

$$k_t^{\text{formal}} = k^{\mathrm{M}}e^{-\delta(t-T)} + \frac{A}{z\delta\left(\delta+\rho\right)}\left(1-e^{-\delta(t-T)}\right)$$

$$k_\infty^{\text{formal}} = \frac{A}{z\delta\left(\delta+\rho\right)} \tag{5}$$

Lifetime profits of the formal firm, discounted back to period zero are

$$\pi^{\text{formal}}\left(k^{\mathrm{M}}\right) = e^{-\rho T}\left(\frac{Ak^{\mathrm{M}}}{\delta+\rho} + \frac{A}{2z\rho\left(\delta+\rho\right)^2} - \frac{rL}{\rho}\right)$$

We can now compare an ever-informal to an ever-formal firm. Because $A > A_j^*$ formal firms invest more than informal firms with the result of a capital stock growing at a faster pace and reaching a higher steady state capital level $k_\infty^{\text{formal}} > k_{j,\infty}^{\text{informal}}$ (this holds for all firms j). In the economy studied all formal firms converge to the same size k_∞^{formal} (whereas in the informal economy we have a distribution of firm sizes).

4.3 Migration

We now consider an informal firm migrating to formality at time T holding a capital stock k^{M}, both of which are still taken as given. We do not yet know if migration will actually be attractive and if so, at what capital level and when it will take place. At this stage we describe the behavior given that such migration takes place.

From the formal firm problem above the continuation value from migrating to formality at a level of capital k^{M} at time T equals $d\pi\left(k^{\mathrm{M}}\right)/dk^{\mathrm{M}} = e^{-\rho T}A/\left(\delta+\rho\right)$. That is; from time T and on the firm is formal and the value of formal firm profits $\pi\left(k^{\mathrm{M}}\right)$, discounted to $t=0$, determines how much the capital stock k^{M} is worth. The migrating firm therefore solves the following problem

$$\text{Choose } i_t \text{ to Max } \int_0^T \left(A_j^*k_t - \frac{z}{2}i_t^2\right)e^{-\rho t}dt \qquad \text{s.t.}$$

$\dot{k}_t = i_t - \delta k_t$ and $k(0) = k_0$ and $\lambda_T = e^{-\rho T}\frac{A}{\delta+\rho}$

λ_t is the Lagrange multiplier of the Hamiltonian; $H = \left(A_j^* k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} + \lambda_t (i_t - \delta k_t)$. The solution to this optimization problem is

$$
\begin{aligned}
i_{j,t}^{\text{migration}} &= \frac{A_j^*}{z(\delta+\rho)} + \frac{A - A_j^*}{z(\delta+\rho)} e^{(\delta+\rho)(t-T)} \\
k_{j,t}^{\text{migration}} &= k_0 e^{-\delta t} + \frac{A_j^*\left(1 - e^{-\delta t}\right)}{z\delta(\delta+\rho)} + \frac{\left(A - A_j^*\right)\left(e^{(\delta+\rho)(t-T)} - e^{-(\delta+\rho)T-\delta t}\right)}{z(\delta+\rho)(2\delta+\rho)}
\end{aligned} \tag{6}
$$

Investments start at the level of investments of the ever-informal firm (the second term in $i_{j,t}^{\text{migration}}$ is roughly zero at $t = 0$). Closer to migration, investments start increasing. They equal the level of formal investments at time T, when the firm becomes formal and starts paying interest payments rL.

This investment behavior means that the migrating firm, instead of withdrawing profits as an informal firm would have done, uses more resources to build its capital stock *in anticipation of migration*. Comparing its capital stock with an informal firm then, for any given time it will be higher. This implies that the migrating firm, before it migrates, accepts to pay higher penalties - these are proportional to the capital stock - while still informal.

5 Determining formalization behavior of a firm

In section 5 we study individual firm investment, growth and migration behavior implied by the dynamic model.

There are two different "evolutions" a firm can go through. Either it is an "ever-informal" firm or it is first a "migration firm" that turns into a "formal firm" at time T. We continue the model description by studying *if* firms will become formal at all and if so at what firm size (k^{M}) and when in time (T).

An additional assumption is used in the analysis that follows: we set the initial capital stock k_0 to zero. The plausibility of this assumption depends on how large the legal cost L is in relation to initial capital. As our objective is to study formalization decisions when L is "large" the initial capital stock does not affect the analysis much[9].

In order to state some general and qualitative conclusions about how firm migration behavior depends on the informal productivity, we continue to view A_j^* as "closed" through this section, without explicitly referring to our micro explanation. That is, for now we continue to view A_j^* as a parameter that varies

[9] Technically, rL being "large" compared to k_0 will mean that $A_j^* k_0 << rL$.

between firms. We will leave this simplified story in the next section where we look at the universe of firms and when we discuss policy.

5.1 Levels of migration

The expressions in the migration subsection above are valid for an informal firm that will become formal at time T. If we set $t = T = \infty$ in (6), we get the capital level that will be accumulated for a firm that migrates in the distant future:

$$k_{j,\infty}^{\text{migration}} = \frac{A_j^* (\delta + \rho) + \delta A}{z\delta (\delta + \rho)(2\delta + \rho)} \tag{7}$$

By comparing the capital levels at infinity for informal, migrating and formal firms we see that $k_{j,\infty}^{\text{informal}} < k_{j,\infty}^{\text{migration}} < k_{\infty}^{\text{formal}}$. Before we derive the optimal level and time at which a firm migrates we can conclude that an informal firm can, by investing more and therefore building a bigger capital stock *in anticipation of migration* to formality, reach a capital level of $k_{j,\infty}^{\text{migration}}$. It will be optimal to follow this capital accumulation path if and only if the level at which to migrate is somewhere below $k_{j,\infty}^{\text{migration}}$. By following this path the informal firm lowers its initial profits, pays higher penalties because of a larger capital stock, but after migration makes enough profits to compensate for the early losses (compared to the ever-informal path). If the migration level of capital cannot be reached the firm will instead find it optimal to be an ever-informal firm and converge to $k_{j,\infty}^{\text{informal}}$.

The level of capital at which a firm chooses to migrate to formality, k^{M}, is determined from the aggregate profit function (informal + formal) of a migrating firm. The marginal value of capital k^{M} is simply $A_j^* k^{\text{M}}$ if the capital is deployed informally. If it is deployed formally it has a marginal value Ak^{M} but a cost of rL is incurred. Investment effects cancel out because at time T they are equal whether the firm is informal or formal (given that migration will occur)[10]. The marginal returns must be equal at the optimum migration level, therefore this optimum migration level equals

$$k_j^{\text{M}} = \frac{rL}{A - A_j^*} \tag{8}$$

where we have added subindex j because each firm has its own optimal migration level. From the derivation of k_j^{M} and by comparing it to the obtainable capital stock $k_{j,\infty}^{\text{migration}}$, we can now state the following two propositions regarding migration levels and behavior of firms.

[10] Formally, differentiate (informal + formal) profits with respect to T and solve for the optimal capital level: $\frac{d}{dT}\left(\int_0^T \left(A_j^* k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt + \int_T^\infty \left(Ak_t - \frac{z}{2} i_t^2 - rL\right) e^{-\rho t} dt\right) = 0$

Proposition 1: The level of capital at which firms migrate is firm specific and equals $k_j^{\mathrm{M}} = rL/(A - A_j^*)$. It *increases* in the productivity of the informal firm.

Higher migration costs rL increase the firm size needed to make it beneficial to become formal. A larger capital stock is needed to make it worthwhile start paying the interest payments rL. Proposition 1 also implies that although the least productive informal firms have a smaller possibility to accumulate capital they also do not need to accumulate that much capital in order to find it worthwhile to migrate; k_j^{M} is smaller for low A_j^*. This effect is highly nonlinear in A_j^*.

Because A_j^* increases less than one to one in A (from (3)) we also see that k_j^{M} is smaller the higher is A which is what is expected: a large productivity difference between formality and informality acts as a strong incentive to become formal.

The next proposition, answering which firms become formal, follows from the condition for migration to take place, i.e. that the obtainable capital stock for a migration firm, $k_{j,\infty}^{\text{migration}}$, be larger than its migration level, k_j^{M}:

$$\frac{A_j^* (\delta + \rho) + \delta A}{z\delta (\delta + \rho)(2\delta + \rho)} \geq \frac{rL}{A - A_j^*} \tag{9}$$

Proposition 2: An informal firm will migrate to formality if and only if its long-run capital stock in anticipation of migration, $k_{j,\infty}^{\text{migration}}$, is larger than its optimal migration level k_j^{M}. All firms that do not fulfill this criterion remain informal forever. Increases in r, L, z, δ and ρ make less firms become formal. Increases in A has the opposite effect. For the informal productivity parameter A_j^* there is an intermediate range for which there is migration. Depending on the range of possible values of A_j^* (which is determined from the penalty policy), formality may thus coexist with both low- and high-end informality.

A higher cost of migration, through r or L, unambiguously increases the level of capital needed for which it is beneficial to migrate. Because each firm has an upper limit in the capital it can accumulate and this does not change with changes in the migration costs, less firms will be able to accumulate a capital stock big enough to become formal. A higher cost z of investing always reduces investment and the capital stock that can be obtained and is therefore also associated with more informality. A higher discount rate ρ and a higher depreciation rate of capital δ also act to reduce investment and the obtainable capital stock and thus produce more informality.

The informal productivity A_j^* plays a crucial role in determining migration behavior of firms. A low-productivity informal firm adds very slowly to its capital stock and it should therefore be more difficult to migrate to formality (LHS in 9). As indicated in proposition 1 however there is a balancing effect through a lower threshold for migration as well (RHS in 9). For low values on the productivity, investments increase linearly with A_j^* but the threshold does not change much when A_j^* increases. We should thus expect possibilities for migration to improve and some lower bound A^{low}, above which firms will actually migrate. The "accumulation effect" thus dominates the "threshold effect" for low but increasing informal productivities. The very most unproductive firms ($A_j^* \leq A^{\text{low}}$) will thus end up in an "informality trap" whereas more productive firms will formalize their businesses.

In the high end of the productivity range the threshold effect will dominate. The threshold for migration is high and increases rapidly with A_j^*, therefore we should expect an upper bound A^{high}, above which high-productivity informal firms do not find migration attractive. There is thus an intermediate productivity range, $A^{\text{low}} \leq A_j^* \leq A^{\text{high}}$, for which it is both feasible and beneficial to pay the fixed cost of migration and formalize the business. There may thus be a low-productivity "informality trap" as well as high-end voluntary informality[11].

In a preview of the effects from the penalty policy it is clear that for any firm in the "left side" in the productivity range higher penalties tend to reinforce the "informality trap". Accumulation of capital is further discouraged, firms invest less and if they migrate at all they do it later. Around the threshold A^{low} a firm that for a certain (lower) penalty accepts to pay a lot of penalties, in anticipation of migration, instead makes a different choice for a slightly higher penalty. Now it no longer wants to become formal, it will not invest and grow as much and it will converge to a smaller firm size.

The story is different for high-productivity informal firms. Higher penalties will act as an incentive to give up current profits and speed up investment in order to be able to pay the formalization fee L. The threshold effect thus dominates for high-productivity firms whereas the accumulation effect dominates for low-productivity firms.

Similar to A_j^*, the formal productivity A has an effect on both the investment behavior and on the threshold itself. These effects act in the same direction; higher formal productivity increases migration investments as well as lowers the threshold. A ceteris paribus increase in A thus widens the range of informal firms that choose to migrate to formality.

[11] The thresholds are obtained by solving for A_j^* in $k_{j,\infty}^{\text{migration}} = k_j^{\text{M}}$;

$A^{\text{low}} = \frac{\rho A - (2\delta+\rho)^{1/2}\left(A^2(2\delta+\rho) - 4z\delta rL(\delta+\rho)^2\right)^{1/2}}{2(\delta+\rho)}$

$A^{\text{high}} = \frac{\rho A + (2\delta+\rho)^{1/2}\left(A^2(2\delta+\rho) - 4z\delta rL(\delta+\rho)^2\right)^{1/2}}{2(\delta+\rho)}$

For $A^{\text{low}} \leq A_j^* \leq A^{\text{high}}$ we have $k_{j,\infty}^{\text{migration}} \geq k_j^{\text{M}}$ and hence migration.

5.2 Time of migration

Whereas propositions 1-2 and the reasoning above determine whether migration takes place we are also interested in when such migration occurs. Proposition 2 only tells us that at some time in the future, possibly very distant, migration occurs. It does not tell us much about the timing of migration and its dependence on the parameters of the problem.

The migration level k_j^{M} and the time path of capital for a migrating firm determine when an informal firm migrates. We let $t = T$ in the capital path for a migrating firm (6) to get the capital level as a function of the migration time T:

$$k_{j,T}^{\text{migration}} = \frac{A_j^* \left(1 - e^{-\delta T}\right)}{z\delta \left(\delta + \rho\right)} + \frac{\left(A - A_j^*\right) \left(1 - e^{-(2\delta+\rho)T}\right)}{z \left(\delta + \rho\right) \left(2\delta + \rho\right)} \tag{10}$$

In principle we can now solve for T from $k_{j,T}^{\text{migration}} = k_j^{\mathrm{M}}$. Although there is no explicit analytical solution for T its dependence on the parameters of the problem can be characterized as in the following proposition;

Proposition 3: The time until migration of a firm, given that it migrates, is increasing in r, L, z, δ and ρ. It is also unambiguously decreasing in A. The migration time as a function of A_j^* is U-shaped; at A_j^* it is infinite, it then decreases to a minimum value and then increases to infinity again at A_j^*

T:s dependence on r, L, z, δ and ρ is straightforward. Higher migration costs through r and L delay migration as do lower investments caused by increases in z, δ and ρ. Higher formal productivity A increases pre-migration investments for any given T and also lowers the migration level, hence makes for earlier migration. With respect to A_j^* the two opposing effects from above; the "accumulation effect" of more investment and the "threshold effect" from less attractive migration determine the behavior once again. T as a function of A_j^* is U-shaped between A^{low} and A^{high}. At both ends of the migration spectrum the informal firm will just reach the migration level k_j^{M} at $T = \infty$. In between these extremes the migration time decreases to a minimum for some intermediate value of A_j^*. The firms that first formalize their business are thus firms with an intermediate productivity. For such firms the accumulation effect and threshold effect combine to produce a fast capital accumulation path without increasing the migration threshold by much.

In a second preview on policy we can thus state that, through the effects that penalties have on productivities A_j^*, it is possible for the policy maker to affect migration times. Penalties producing intermediate values on A_j^*, i.e. somewhere between A^{low} and A^{high}, make for fast migration.

5.3 Firm behavior over time

We let two graphs illustrate how a migrating firm grows over time. The firm will invest according to the migration section above, starting at the level of investments of the informal firm and then gradually increase investments up to the level of investments of the formal firm i^{formal} which it reaches at time T. The firm pays L to become formal, financed from borrowing, and starts paying interest payments rL. The capital stock of the firm will then grow until it reaches $k_{\infty}^{\text{formal}}$ where investment just replace the depreciated capital each period.

The picture below shows, for a certain choice of parameters, the development of investments and of the capital stock of a migrating firm, i.e. a firm for which $k_{j,\infty}^{\text{migration}}$, the upper solid line in the right graph, is larger than k_j^{M} (dotted line in the right graph). Investments increase from i_j^{informal} to i^{formal} and capital converges to $k_{\infty}^{\text{formal}}$. The subindex j has been omitted in the graphs.

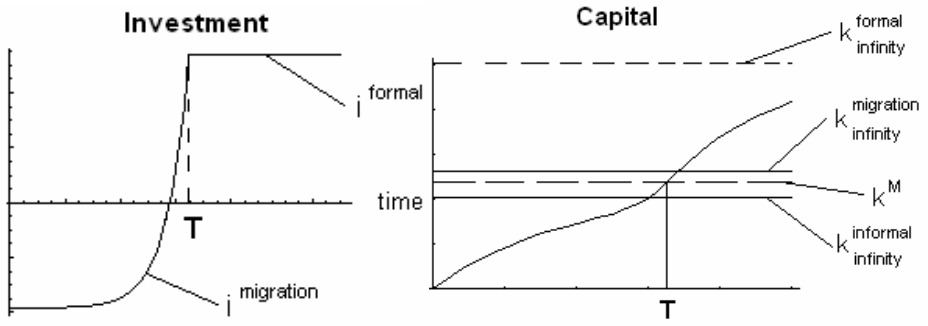

Figure 2

If the condition for migration had not been fulfilled, for instance due to a higher L pushing k_j^{M} above $k_{j,\infty}^{\text{migration}}$ (upper solid line), investment would have been constant at the starting level and the capital stock would have converged at $k_{j,\infty}^{\text{informal}}$, the lower solid line.

6 Aggregate behavior

In section 6 we study aggregate firm investment, growth and migration behavior implied by the dynamic model.

From the analysis above we learn that there is an intermediate productivity range in which it is possible and desirable for firms to become formal. Outside of this interval it is either not possible (low values of A_j^*) or not desirable (high values of A_j^*) to formalize the business. The thresholds for migration, A^{low} and A^{high}, repeated below for convenience, are determined by the exogenous parameters of the model:

$$A^{\text{low}} = \frac{\rho A - (2\delta+\rho)^{1/2}\left(A^2(2\delta+\rho) - 4z\delta rL(\delta+\rho)^2\right)^{1/2}}{2(\delta+\rho)}$$
$$A^{\text{high}} = \frac{\rho A + (2\delta+\rho)^{1/2}\left(A^2(2\delta+\rho) - 4z\delta rL(\delta+\rho)^2\right)^{1/2}}{2(\delta+\rho)}$$

The comparative statics of these expressions are straightforward. Higher values of the migration cost rL make migration more costly and narrows the gap between A^{low} and A^{high} and there is a cutoff migration level $(rL)_{\max}$ above which no firm can become formal. Higher values of the investment cost z, the depreciation rate δ and the time discount factor ρ all have a negative impact on investment and make less firms migrate. The productivity of the firm once formal, A, widens the migration gap for two reasons. First, the incentive to become formal is strengthened by a higher formal productivity and, second, investment of the migrating firm increases in A.

In order to study aggregate migration behavior, i.e. what firms migrate to formality, we need to connect our micro analysis in section 3 with the dynamic analysis. From the discussion in section 3 we know that each firm α_j has its own unique productivity and the productivity range will be determined by the penalty rate x and the institutional parameter e.

In order to get an idea of how the economy may evolve over time, and before discussing policy, assume that $x = A$ and $e \to 0$ such that the productivity range goes from 0 to A as in section 3.3.1. Also assume that rL is below $(rL)_{\max}$ so that we have some migration. There will be a range of α:s, α^{low} to α^{high}, corresponding to A^{low} to A^{high}, for which there is migration.

Firms have zero initial capital and therefore the aggregate economy also starts at zero. The first four graphs below show the capital stock as a function of the informal productivity at four different times (thick line). The criterion for migration, $k_{j,T}^{\text{migration}} \geq k_j^{\text{M}}$, has been evaluated and the k_j^{M}-curve between α^{low} and α^{high} is shown in the graphs, together with the curve for long-term informal firm size, $k_\infty^{\text{informal}}$ (the subindex j has been omitted in the graphs.). It is when a firm with a productivity between α^{low} and α^{high} reaches its own k^{M}-level that it migrates.

Until the first firm migrates to formality, something which happens for an intermediate productivity firm, *all* firms are informal (the two upper graphs). In the intermediate range of productivities between α^{low} and α^{high} the capital accumulation path corresponds with the migration investment function outlined above. Such firms thus forsake current profits in order to reach a high enough capital stock for migration. The "bulb" on the curves corresponds to the larger capital accumulation for these future migrating firms. The upper right graph corresponds to a situation where the first informal firm is about to become formal (having reached its migration level k_j^{M}).

In the lower left graph some firms have migrated and continue to accumulate capital, now investing always i^{formal}, growing towards the long-run formal firm size $k_{\infty}^{\text{formal}}$. In the lower right graph, showing the situation at $t = \infty$, the informal and formal firms have converged to their long-term sizes.

In both ends of the α_j-spectrum firms invest according to the ever-informal case and have a slower capital growth compared to the case had they been migrators. They approach their individual $k_{\infty}^{\text{informal}}$-level. For the parametrization given, and as a preview to what we can expect from high penalties in a low-enforcement environment, we have a low-end "informality trap", a middle range of large formal firms and a high-productivity range of informal firms.

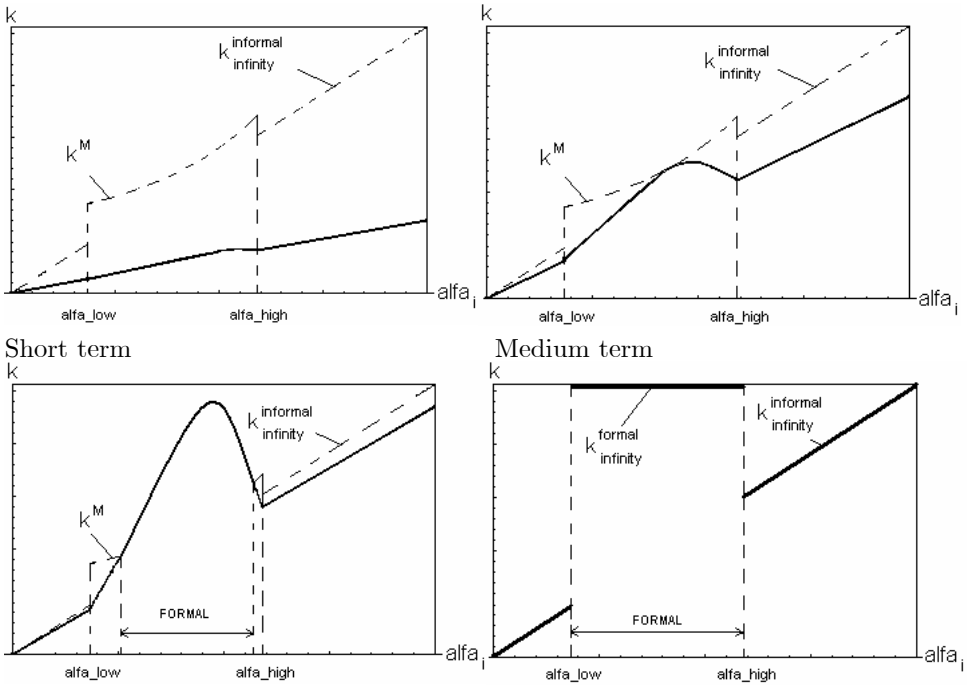

Figure 3

Figure 4 shows the growth of the total capital stock of the economy over time for the same parametrization that has produced the four preceding graphs. Although formality increases rapidly after the first migration, for this parametrization a substantial fraction of the overall capital stock is informal at all times.

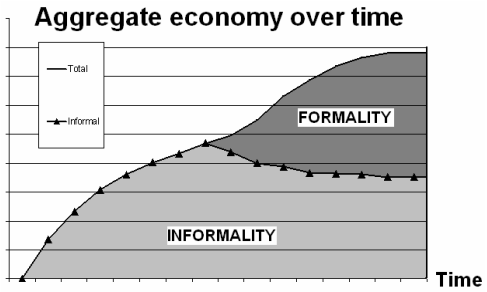

Figure 4

7 Policy

We now study the effects of penalty policy by first introducing a benevolent government that has a negative externality from informality in its welfare function. For simplicity, first assume that the government only values formal but not at all informal firm profits and that the externality from informality is proportional to the informal capital stock. The larger the informal firms are the larger is the negative externality. This may be due to lack of compliance with environmental and health regulation on part of the informal firms. Further assume that the government sets a penalty in period zero and then sticks to this policy. We look at the two different institutional environments, $e \to 0$ and $e = 1/2$, in turn and we assume that parameters are such as to make the problem interesting, i.e. that the migration cost rL is neither too low nor too high producing a migration interval $0 \leq A^{\text{low}} \leq A^{\text{high}} \leq A$.

7.1 Low e - benevolent government

In a typical developing country it is unlikely that the government can fully enforce a penalty policy vis-a-vis informal firms. In the context of our model, e is low. In this case, displayed in section 3.3.1 above, for positive penalties there is large variation in productivities between firms. There is heterogeneity in the skill of dealing with informality, i.e. in α_j, and this difference in skills is fully reflected in the productivity; $A_j^* = A - (1 - \alpha_j)\, x$.

As long as penalties are small, such that the minimum productivity is above A^{high}, there is no migration to formality at all, clearly not optimal in this case. If the externality from informality is not too large the optimum penalty will be to set x such that the lowest productivity firm ($\alpha_j = 0$) gets productivity A^{low}, i.e. $x = A - A^{\text{low}}$ as in the left graph below. In this case we achieve as much migration to formality as possible although there is still substantial high-end informality. Therefore, with a larger externality from informal production the

government may prefer to set a penalty all the way up to $x = A$ as in the right graph[12].

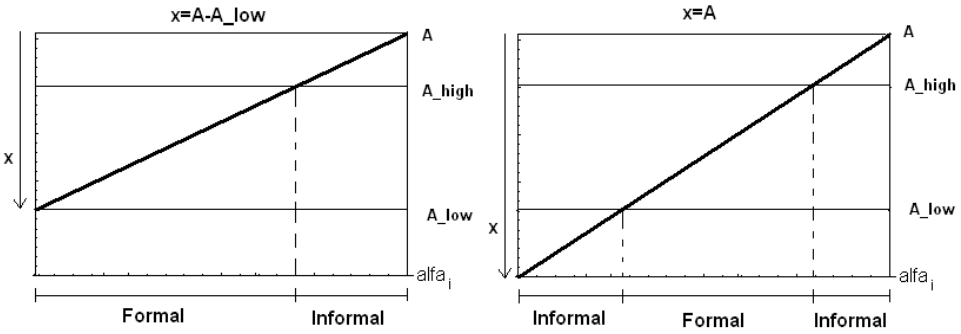

Figure 5

The analysis implies that a benevolent government, with a large externality from informality in its welfare function, can set high penalties that create a lot of low-end informality. The firms with the lowest ability in dealing with being informal end up investing very little and are unable to become formal because of the accumulation effect.[13]. We get a long-term distribution of firm sizes looking as in the lower right graph of figure 3, with a substantial fraction of very small firms, a "missing middle" in firm sizes, larger informal firms and also a portion of equally sized large formal firms.

The analysis carries over to the case where the government also values informal profits, although the externality from informality must be higher for the government to choose the maximum penalty.

Assume that the government instead wants to produce fast migration. Also in this case will it be unable to fully achieve its goal, as many firms can fully shed themselves from penalties.

7.2 High e - benevolent government

Different from the institutional context above it may instead be the case that the implementation of policies is easier; e is high. In our alternative parametrization, i.e. $e = 1/2$, the government can go a long way in affecting the investment incentives of firms without creating low-end informality. If the productivity

[12] We should remember that we have limited the maximum x to equal A in order to avoid negative production. However, it dos not seem unlikely that in reality penalties can be higher than current production possibilities which clearly would create more low-end informality.

[13] It does not matter for the analysis presented whether the government only cares about the long-term distribution of firm sizes or the profits during migration

thresholds continue looking as in figure 5 (the migration interval is big enough) the optimal policy will be to eliminate high-end informality, ending up with a situation as in figure 6. The reason to set a somewhat higher penalty rate that the minimum necessary to assure all firms migrate is to assure faster migration and therefore a higher present value of formal profits (and hence shorter time with externality from informality). Indeed, the better enforcement there is, the flatter the productivity of informal firms and the better the government can "fine tune" its policy.

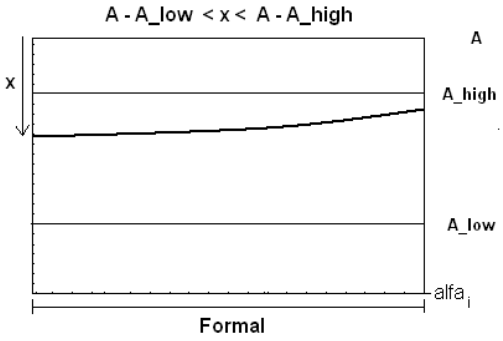

Figure 6

As for the policy analysis so far we can conclude that the benevolent government will be better at achieving its objective the better enforcement there is. Vice versa, settings in which there is much leeway for individual behavior on part of firms may make the government set high penalties. In a high-e environment it is easier for the government to achieve its objectives, fine-tuning the "carrot and stick" penalty tool.

7.3 The malevolent case

The characterization of the penalty policy as being the result of a benevolent government that exercises some control over informal firms is likely to be far from reality in some countries. It may be more likely that small informal firms face low-level government agents that in theory should assure that firms comply with rules and regulations but in practice take bribes against letting firms operate clandestinely. In this case an interpretation of e is that with low values the corrupt bureaucrat or police is not very efficient in doing what it is doing and simply cracks down on the most visible firms. With a high e we can instead think of such a corrupt bureaucracy as being highly efficient.

7.3.1 Static case

First assume that such a bribe-taker simply maximizes period 0 revenue (all firms informal) without taking into account the effect of penalties on future period capital stocks. In this case the "optimal penalty" will clearly be $x = A$. The bureaucrat simply confiscates as much as possible. Interpreting the bureaucrat as "weak" ($e \to 0$) this will yield an informality/formality division as in figure 5B above. Instead, a corrupt bureaucrat that is better at enforcing bribe payments ($e = 1/2$) will be much more detrimental to small informal firms, as the graph below shows. Low-end informality is abundant and only the highly able firms will be able to migrate to formality.

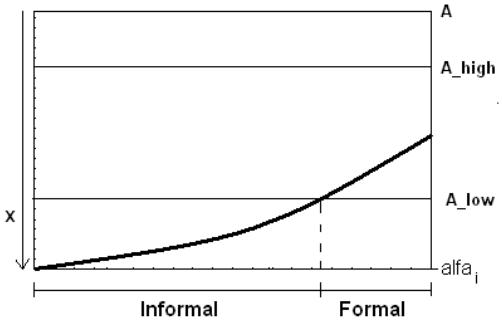

Figure 7

7.3.2 Dynamic case

Assume instead that the bureaucrat sets a penalty and commits to it also in future periods, taking into account the growth of the capital stock in setting the penalty. Assume that migration is so far in the future and/or the bureaucrat's discount rate so high that the he does not take into account the possible future migration of the middle range firms. It considers all firms "ever informal" with capital stocks growing according to $k_{j,t}^{\text{informal}}$ in (4), i.e. proportional to A_j^*, and hence chooses x to maximize $\int_0^{\alpha_{\max}} xA_j^* d\alpha_j$. We get the following "optimal" penalties for the two cases:

$e \to 0$:

Choose x to Max $\int_0^1 x\left(A - (1-\alpha_j)\,x\right) d\alpha_j \qquad \Rightarrow \qquad x = A$

$e = 1/2$:

$$\text{Choose } x \text{ to Max } \int_0^{\sqrt{2}} x\left(A - x + \frac{\alpha_j^2 x^2}{4A}\right) d\alpha_j \qquad \Rightarrow \qquad x = \frac{2A}{\sqrt{2}+2} \approx 0.6A$$

Because firms in the low-e environment are better at shedding themselves from the bribes collected the future capital stocks are not affected as much and the bureaucrat can therefore set a higher penalty than in the case where it is harder to escape the bureaucrat. The firms best at shedding themselves from penalties will be able to build large capital stocks and therefore constitute a source of future revenue for the bureaucrat. The low-productivity firms will suffer a lot however. In the high-enforcement scenario instead penalties will be lower because the future capital stock of the largest informal firms are affected more.

From the penalty policy analysis above it is clear that the highest penalties will be set in environments with little enforcement and by malevolent bureaucrats. A benevolent government may also end up setting high penalties in an environment where firms can shed themselves. In high-enforcement environment, a benevolent government will set lower penalties.

7.4 Other policy questions

7.4.1 Taxation and public goods

Lack of incentives in the formal sector itself is often claimed to be the main determinant of informality. As this paper has argued, large entry costs into formality together with harsh penalties can create both an informality trap and voluntary high-end informality. For low-productivity firms the threshold effect dominates. These firms simply cannot accumulate enough capital for migration to be possible. For a middle range of firms it is both possible and beneficial to become formal whereas for the high productivity informal firms migration is not attractive.

Consider taxation and public goods in the formal production function, such that the productivity parameter is now A^{formal}, rather than A, once a firm has become formal. If the net effect of taxation and public goods provision on formal productivity is positive, such that $A^{\text{formal}} > A$ there will be an unambiguous effect that increases migration. It has two components, as seen in (9), and as discussed in section 5. First, the threshold for migration shifts down. It has become more attractive to formalize the business, the informal firms therefore become formal at a smaller size. Second, investment for the migrating firms, in anticipation of becoming formal, increase. This is because the profits once

formal, through the continuation value in the migration optimization problem, has increased and investment in the capital stock becomes more attractive.

It should be noted that, for penalties $x = 0$ all informal firms are equal and it may be that, although $A^{\text{formal}} > A$, the informal productivity A may fall outside the range of migration produced by A^{formal}. All firms are "high-end" informals in the absence of penalties. In this case there is a clear role for a "carrot and stick"-policy. Introducing $x > 0$ will cause a "threshold" effect that make high-end informal firms opt for formality.

7.4.2 Entry costs

As has been widely documented and discussed in this paper, entry costs to formality can be very high. A main purpose of the work has been to study what effects such high entry costs have on firm behavior, especially investment and growth. The existence of high entry barriers may be due to a multitude of reasons; "insiders" (formals) protecting themselves from "outside" competition by lobbying the government for protection, a source of revenue for low-level government bureaucrats or, indeed, the entry procedures may serve a useful purpose such as health regulations for firms in chemical industry. We have taken the view that, at least in the short run, it is so difficult and costly to reform the "entry procedures" that they can be considered fixed. Furthermore, even if the rules are not completely fixed in reality, it may take a very long time for such knowledge to affect firm behavior as firms from past experiences expect entry procedures to remain cumbersome.

8 Discussion

A dynamic model of capital accumulation has been used to study the possible migration of informal firms to formality. The dynamic aspect is crucial; only over time can small informal firms be expected to reach the size necessary for formalization to be feasible. Furthermore, in a dynamic setting, not only a migration threshold effect but also a capital accumulation effect determine behavior of firms.

Depending on how policy vis-a-vis small informal firms is viewed, the policy outcome can range from large parts of firms ending up in an informality trap with little investment and small firm sizes to a government that can fine tune policies to assure rapid formalization of firms. The institutional context however, here modeled as a parameter determining how easy it is for firms to escape the authorities, puts a bound on how effective the policy maker can be in achieving its objectives.

9 References

Antunes, A. R., de V. Cavalcanti, T. V, 2006. Start-up costs, limited enforcement and the hidden economy. European Economic Review 51, 203-224
Banerjee, A. B., Newman, A. F., 1994. Poverty, Incentives and Development, American Economic Review 84, 211-15
De Soto, H., 1989. The Other Path. New York: Harper and Row.
Doingbusiness.org, 2007
Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The regulation of entry. The Quarterly Journal of Economics 117, 1-37.
Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. The Review of Economic Studies 60, 35-52
Loayza, N. V., 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
Prado Jr., J. M., 2007. Government Policy in the Formal and Informal Sectors, IIES/Seminar Papers #751.
Rausch, J. E.,1991. Modelling the informal sector informally. Journal of Development Economics 35, 33-47
Schneider, F., 2002. Size and measurement of the informal economy in 110 countries, Workshop on Australian National tax centre
SEBRAE, 2007. The Brazilian small enterprise authority. www.sebrae.gov.br
Stone, Levy, Paredes, 1996. Public Institutions and Private Transactions: A Comparative Analysis of the Legal and Regulatory Environment for Business Transactions in Brazil and Chile, in "Empirical studies in institutional change" (Alston, Eggertsson & North eds.)
Tokman (1992), The informal economy in Latin America