Informal firms, investment incentives and formalization*

Anders Fredriksson$^{a,\dagger}$

$^a$Institute for International Economic Studies (IIES),
Stockholm University, SE-106 91 Stockholm, Sweden

April 20, 2009

Abstract

In a typical developing country, the majority of small firms are informal and entry costs into formality are high. This paper is motivated by these two observations. It asks the question what can be expected in terms of firm investment, growth and formalization in such a setting. It also studies the effects of policies towards the informal sector, on formalization decisions.

Several interesting results emerge from the analysis of the dynamic model set up to study these issues. First, the investment paths and growth trajectories differ substantially between firms that choose to formalize and those (ex-ante almost identical firms) that do not. Second, the formalization decision depends non-trivially on the productivity of the informal firm, due to the balancing of an accumulation effect and a threshold effect. This in turn has an effect on how a policy to incentivize informal firms to become formal should be designed. Third, when aggregating over firms, the long-run firm size distribution exhibits a range of small firms and a range of larger firms, but also a "missing middle", much in line with actual firm size distributions observed in developing countries. Fourth, the long-run firm-size distribution turns out to depend on the initial firm level stock of capital, a result that can be interpreted as a poverty/informality trap.

JEL classification: E22, E26, O10, O17

Keywords: Informal firms, Investment, Entry costs, Non-convexities, Formalization

*I thank Shon Ferguson, Harry Flam, Gustav Hansson, John Hassler, Per Krusell, Ted Miguel, José Mauricio Prado, Jesper Stage, Jakob Svensson and Fabrizio Zilibotti for helpful discussions, feedback and suggestions. I also thank workshop participants at the Department of Economics at Stockholm University, the Berkeley development lunch, IIES brown bag, and participants at SUDSWEc and at the Nordic Conference of Development Economics in Copenhagen, both held in 2007. All errors are mine.

$^\dagger$Tel.: +46-8-162326; fax: +46-8-161443. E-mail address: anders.fredriksson@iies.su.se.
1 Introduction

In a typical developing country, the majority of small firms are informal and entry costs into formality are high. This paper is motivated by these two observations. It asks the question what can be expected in terms of firm investment, growth and formalization in such a setting. The paper focuses on firms' incentive to invest when, at some future point in time, large entry costs into formality have to be paid. The effect from a penalty policy, on informal firm investment, growth and formalization is also discussed.

The observation that most small firms in developing countries are informal is well-established. A recent enterprise survey in Brazil shows that 90% of the smallest firms, i.e. of firms with 1-5 employees, have not gone through the procedure to register as a legal entity (SEBRAE, 2005). An enterprise survey in Mexico, the other large Latin American economy, shows similar values (INEGI, 2003). Studies and accounts from other developing countries indicate similar degrees of informality among the smallest firms in the economy (see for instance Bigsten et al., 2004, for Kenya and de Soto, 1989, for Peru).

Turning to bureaucratic and legal costs facing small and medium enterprises, such costs have received considerable attention in recent development research. In particular, the work by de Soto (1989) and Djankov et al. (2002) has directed attention to substantial government-related costs of "doing business" and entry into formality. Examples of such costs are start-up fees, financial costs incurred in order to pay taxes (except for the taxes themselves), financial costs related to hiring and laying off workers, as well as the time spent with these activities. The costs can be substantial. Whereas it costs USD 370 to start a firm in the US, the average cost in Latin America is around USD 1240, as reported by the World Bank Doing Business project. Average monthly income per capita was USD 3840 in the United States in 2007, meaning that 3 days of work generates an income equal to the firm start-up cost. In Latin America, average monthly income was one tenth as much, or USD 380. It thus takes more than 3 months of work to generate an income equal to the firm start-up cost. Furthermore, income levels in the informal sector in Latin America are typically much lower than the official GNI figures, meaning that it takes even longer to generate an income equal to the firm start up costs.

As implied by the above, the definition of an informal firm used in this paper is a firm that has not gone through the registration procedure at the government bureaucracy.

The combination of small informal firms and large formalization costs has motivated setting up a simple dynamic model of profit maximizing firms. Firms can invest in their capital stock, grow larger, and, possibly over time, become

\footnote{Firm start up costs and income data come from World Bank (2009A) and World Bank (2009B), respectively. The numbers concern the year 2007. See section 3 for details.}
formal. The cost to become formal is taken literally: at one instant in time the firm can choose to pay the formalization fee, defined as $F$, a fee that represents all the costs to register the firm at the government bureaucracy. Having paid $F$, the firm changes status from informal to formal and gets a productivity benefit.

How do formalization costs, to be paid at some future date, affect investment today? At what firm size and when do firms choose, if at all, to become formal? What are the crucial parameters affecting firm formalization? What is the effect from credit constraints on the formalization decision? Can formalization costs lead to poverty traps? How should policy vis-a-vis informal firms be viewed? How can the government affect the formalization decision and what is an appropriate government policy vis-a-vis informal firms? These questions are addressed in this paper.

Several interesting results emerge from the analysis of the tractable dynamic model. First, the investment paths and growth trajectories differ substantially between firms that choose to formalize and those (ex-ante almost identical firms) that do not. Second, the formalization decision depends non-trivially on the productivity of the informal firm, due to the balancing of an accumulation effect and a threshold effect. This in turn has an effect on how policy designed to incentivize informal firms to become formal should be designed. Third, when aggregating over firms, the long-run firm size distribution exhibits a range of small firms and a range of larger firms but also a "missing middle", much in line with actual firm size distributions observed in developing countries (Bigsten et al. 2004, Tybout, 2000). Fourth, the long-run firm-size distribution turns out to depend on the initial firm level stock of capital, a result that can be interpreted as a poverty/informality trap.

The paper proceeds as follows: In section 2, the literature to which this paper relates is reviewed and the model to be presented is motivated in relation to this earlier writing. Section 3 discusses formalization costs in different countries and presents some data on income levels in the informal economy, together with typical informal firm capital stocks and profit levels from three recent studies. The dynamic model of firm investment and formalization is presented in section 4 and analyzed in section 5. Some extensions to the analysis, focusing on how the investment and formalization behavior changes when basic assumptions are altered, are in section 6. Section 7 concludes. The appendix presents some of the details in deriving the analytical results.

2 Literature review

An important debate in the literature on the informal sector, preceding the analysis in this paper, is whether small informal entrepreneurial activities at all should be considered proper "firms", or merely temporary subsistence labor
while waiting for a formal job. In early writing on how the economy develops from traditional to modern, Lewis (1954), Todaro (1969) and Harris and Todaro (1971) considered the "urban traditional" sector a source of labor supply for the "modern" sector\(^2\). In none of these papers is the urban traditional sector seen as an important element of economic activity nor a contributor to capital accumulation. It is rather considered a temporary low-productivity subsistence activity.

The entrepreneurial view, that informal small-scale economic activities should be considered entrepreneurs/firms, rather than subsistence activities, has been popularized by de Soto (1989). A change in terminology and focus however, to an informal - rather than an "urban traditional" - sector, stressing entrepreneurship and not only surplus labor, emerged with the writings of the International Labor Organization (Hart, 1973; ILO, 1972). The informal sector/informal economy started to be seen more as a permanent, increasing and diverse phenomenon - "from marginal operations to large enterprises" (Hart, p. 68)\(^3\).

In an "occupational choice" model in the spirit of Lucas (1978), with economic agents differing in entrepreneurial ability, Rauch (1991) studies the choice between being a worker, being an informal entrepreneur or a formal entrepreneur. Both types of entrepreneurs employ workers. The informal entrepreneurial sector arises as a result of a government (above-market clearing) minimum wage policy. The static general equilibrium model delivers predictions on the relative size of the informal sector, firm size distribution, and changes to these from the minimum wage level\(^4\).

Minimum wages that do not clear the market is an example of a government intervention that may lead to an informal sector. The focus here is instead on the effects from government-imposed formalization costs. The paper takes as given the de Soto entrepreneurial view and studies investment and formalization decisions of profit maximizing informal firms.

\(^2\)Starting with the work of Lewis (1954), the traditional urban sector plus rural-to-urban migrants were seen as a source of unlimited labor supply from which the modern sector could get labor at subsistence pay. Todaro (1969) modelled the rural--to urban migration decision, taking into account the existence of an unemployed or underemployed pool of urban traditional workers that competes for the same jobs as the rural migrants. Harris and Todaro (1971) studied a minimum wage policy in a similar setting.

\(^3\)Rauch (1991), Chen (2004) and de Mel et al. (2008) all discuss early writings on the informal economy.

\(^4\)Rauch’s paper can be seen as combining the two views above on informal activity. In recent empirical work from Sri Lanka, de Mel et al. (2008) collect data on personal characteristics from wage workers, own-account workers and from owners of enterprises with 5-50 employees, in order to address the question whether own-account workers resemble wage workers or firm-owners more. By using a "species classification" approach from biology, they classify around 70% of the own-account workers in their study as wage workers and 30% as small- and medium enterprise owners. On the other hand, the authors argue that, given the large number of own-account workers in low-income countries, the possibilities for job creation and growth in this sector should not be ignored. In addition, by exploiting the panel structure of the data on own-account workers, for the fraction of this group that resemble enterprise owners, the growth rate implied by these own-account workers’ transition into being employers is much larger than growth rates found in a comparable study with data from the United States.
The question whether a firm formalizes or not in the face of large such costs involves at least two issues: the formalization costs themselves and the potential gains from formalization. In addition, a modelling choice has to be made. In order to capture effects from large formalization costs on small informal firms, a dynamic framework is appropriate: firms have to grow to a certain size to become formal. A dynamic model can also shed light on how the investment incentives and the resulting growth path today are affected by a future "non-convexity" in the production function.

To the best of my knowledge, this paper is the first to explicitly focus on the investment incentives in anticipation of a formalization cost. The model is similar in spirit however to the literature on non-convexities and poverty traps, a literature that typically focuses on whether initial (wealth) conditions matter for long run allocations.

One basic insight from neoclassical theory is that non-convexities alone will not affect long run allocations. Economic agents could simply borrow to overcome such hurdles. The analysis of models with non-convexities is therefore intimately connected with introducing some other constraint relevant for developing economies, in particular capital market imperfections, which may make individuals or firms unable to converge to a common long run steady state or balanced growth path (Banerjee, 2001 and McKenzie and Woodruff, 2006 discuss this point). The effects from initial capital and credit constraints, on the possibilities for firm formalization, are discussed in this paper.

Typically, the interplay between non-convexities, credit constraints and initial wealth is studied in dynamic occupational choice models with an OLG-structure, in which one generation bequests wealth to the next generation and where individuals have a "warm glow" utility function. In the baseline human capital investment model of Galor and Zeira (1993), this results in a direct relationship between the initial wealth of one generation of a dynasty and the long-run steady state of the same dynasty. The decision of one generation to invest or not invest in human capital results in the same decision being made in the next

5Regarding terminology, the present paper discusses government imposed formalization costs, in the form of going through a firm registration procedure, as the fundamental non-convexity which matters for firm growth. This is different from occupational choice models, such as Banerjee and Newman (1993), Ghatak and Jiang (2002) and Buera (2008), in which the non-convexity is typically a minimum scale investment. The two different types of entry costs may well operate on different levels of firm size: an individual considering starting a manufacturing "firm" may consider buying a machine ("entry"). After having grown, such an informal manufacturing firm, with an established operation and possibly with a number of employees, may consider "formalization". In a recent empirical paper on the return to capital investment for small firms in Mexico, McKenzie and Woodruff (2006), in line with other papers, find high returns on small investments for the smallest firms and thus find no evidence on "entry nonconvexities". They do find lower returns for firms with a capital stock in the USD 1000-2000 range however, and cannot reject that there is a threshold effect "whereby fiscal and bureaucratic costs are faced only by firms above a minimum size" (McKenzie and Woodruff, 2006, page 5).
generation, and so on\textsuperscript{6}. There is no intergenerational saving/investment where a current generation takes into consideration the possibility that the decision of a future generation may be affected by today’s choice.

The framework in this paper is different. Firms maximize profits over the entire life span of the firm. This means that the investment decision is truly intertemporal. The firm considers whether it should build up a capital stock over time, although this may imply current losses, in order to formalize at some later point in time\textsuperscript{7}.

A feature of the present model, differently from most other papers, is that it is possible to solve analytically for the shape of the investment function over time. The comparative statics of the model can therefore be analyzed in a straightforward way.

Turning to the second issue, what is to be gained from formalization? This paper assumes that there is a productivity gain from becoming formal and focuses on the resulting effect on investment incentives while informal, but does not provide one specific channel by which formal productivity is higher.

A non-exhaustive list of aspects that differ between informal and formal firms, from the development literature, includes access to credit and capital, taxes, public goods provisioning, access to risk pooling mechanisms, security in business environment, property rights, marketing possibilities, access to export markets, supplier-buyer relationships and other contracting issues (see for instance de Soto, 1989; Tokman, 1992; Levenson and Maloney, 1998; Bigsten et al. 2004; Chen, 2004; Maloney, 2004)\textsuperscript{8} \textsuperscript{9}.

One mechanism, out of many possible, that affects the productivity of informal firms, and therefore the incentive to become formal, is instead proposed: Penalties and enforcement vis-a-vis informal firms make these firms divert time from production, with lower total production as a result. 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

---

\textsuperscript{6}In an extension, as well as in the occupational choice models of Banerjee and Newman (1993) and Ghatak and Jiang (2002), the entire wealth distribution endogenously determines occupational choices and wages, which in turn affects the bequest to the next generation and the long run equilibrium.

\textsuperscript{7}See Banerjee (2001 pp. 31-32) for a discussion of "joy of giving" vs. "Barro preferences". In an appendix, Galor and Zeira (1993) point out how a "utility of offspring"-approach would affect their results and show that a poverty trap would still result. The occupational choice model of Buera (2008) also uses fully intertemporal preferences.

\textsuperscript{8}The effects from taxation in the formal sector and from differences in public goods provisioning between the sectors are modelled by Loayza, 1995 and Garcia Penalosa and Turnovsky, 2005. Differences in access to outside finance is modelled by Antunes and Cavalcanti, 2007.

\textsuperscript{9}The assumption that formality brings a productivity benefit is not uncontroversial. As an example, much of the discussion in Brazil is centered around (too high) taxation in the formal sector. This paper assumes that formality is desirable, although the framework could in principle allow for firms that do not desire formality.
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, and so on. The set-up, where firms respond to penalties by diverting time from production, allows us to explicitly study the effect of changes in policy, i.e. penalties, on informal firm investments and decisions to formalize.

The main focus of the paper is to study the investment incentives of an individual firm. However, the aggregate formalization behavior of heterogenous firms - differing in an ability parameter (or in initial capital), is also studied. The aim is not to provide an industry evolution model, as in Jovanovic (1982), Hopenhayn (1992) and Melitz (2003). It is rather to display the implications of the non-convexity on long-run firm sizes and formality status, when firms differ in ability and initial capital. These "aggregate" predictions of the model are outlined in section 5.

3 Formalization costs

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 such costs is most likely the "Doing Business" project financed by the World Bank. This data set on costs to start a firm originally covered 75 countries (Djankov et al., 2002), now 181 countries are included (World Bank, 2009A). A summary of the most recent data, from 2009, 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 Gross National Income (GNI). The financial cost to start a business (column 3) is at least 30% of yearly GNI per capita in most of the developing world, and as much as 111% in Sub-Saharan Africa.
Table 1. Number of procedures, duration and cost to register a business in different parts of the world. Source: World Bank, 2009A.

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of procedures</th>
<th>Time (days)</th>
<th>Cost to start a firm / (GNI/capita)</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia &amp; Pacific</td>
<td>8.6</td>
<td>44.2</td>
<td>32.3</td>
</tr>
<tr>
<td>Eastern Europe &amp; Central Asia</td>
<td>7.7</td>
<td>22.6</td>
<td>8.6</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>9.7</td>
<td>64.5</td>
<td>39.1</td>
</tr>
<tr>
<td>Middle East &amp; North Africa</td>
<td>8.4</td>
<td>23.5</td>
<td>41.0</td>
</tr>
<tr>
<td>South Asia</td>
<td>7.4</td>
<td>32.5</td>
<td>31.9</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>10.2</td>
<td>47.8</td>
<td>111.2</td>
</tr>
<tr>
<td>OECD</td>
<td>5.8</td>
<td>13.4</td>
<td>4.9</td>
</tr>
<tr>
<td>United States</td>
<td>6</td>
<td>6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 2 presents data for the year 2007 for the Latin American countries present in the World Bank data, augmented with informal economy income figures from Schneider (2002). Columns 1-3 show that 6-17 different bureaucratic procedures with a total cost of 585-2820 USD and consuming 19-152 days have to be taken in order to formalize a firm. The average is 12 procedures, 58 days and 1238 USD in official cost. All Latin American countries have higher firm start-up cost than the United States and the average cost is 336% of the US cost.10

Column 4 shows the official 2007 GNI/capita figures from World Bank (2009B), column 5 shows the informal GNP/capita figures from Schneider (2002) and columns 6-7 show the ratio between the cost to start a firm to the monthly informal GNP/capita and the ratio between a "total cost" to the monthly informal GNP/capita, respectively11, 12. 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 firm start-up cost and total cost, respectively.

If we focus only on the official cost to start a firm (column 6), then Brazil, the most favorable country, requires 3 months of work to generate the income.
required for the formalization cost. The Latin-American average is 11 times informal GNP/capita and Bolivia and Nicaragua have very high costs in terms of informal income. These costs are high and likely to be prohibitive for many small informal firms.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of procedures</th>
<th>Time (days)</th>
<th>Cost to start a firm (USD)</th>
<th>Official monthly GNI (USD)</th>
<th>Informal monthly GNP (USD)</th>
<th>Start up cost/ (Informal monthly GNP)</th>
<th>Total cost/ (Informal monthly GNP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>14</td>
<td>31</td>
<td>702</td>
<td>504</td>
<td>128</td>
<td>6,5</td>
<td>6,1</td>
</tr>
<tr>
<td>Bolivia</td>
<td>15</td>
<td>50</td>
<td>1891</td>
<td>105</td>
<td>70</td>
<td>26,8</td>
<td>27,6</td>
</tr>
<tr>
<td>Brazil</td>
<td>17</td>
<td>152</td>
<td>585</td>
<td>493</td>
<td>196</td>
<td>3,0</td>
<td>4,7</td>
</tr>
<tr>
<td>Chile</td>
<td>9</td>
<td>27</td>
<td>818</td>
<td>696</td>
<td>138</td>
<td>5,9</td>
<td>6,4</td>
</tr>
<tr>
<td>Colombia</td>
<td>13</td>
<td>44</td>
<td>644</td>
<td>271</td>
<td>106</td>
<td>6,1</td>
<td>6,8</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>12</td>
<td>77</td>
<td>1279</td>
<td>463</td>
<td>121</td>
<td>10,5</td>
<td>11,5</td>
</tr>
<tr>
<td>Dom. Rep.</td>
<td>9</td>
<td>72</td>
<td>1072</td>
<td>296</td>
<td>95</td>
<td>11,3</td>
<td>12,1</td>
</tr>
<tr>
<td>Ecuador</td>
<td>14</td>
<td>65</td>
<td>979</td>
<td>257</td>
<td>88</td>
<td>11,1</td>
<td>12,0</td>
</tr>
<tr>
<td>Guatemala</td>
<td>13</td>
<td>30</td>
<td>1271</td>
<td>203</td>
<td>105</td>
<td>12,1</td>
<td>12,7</td>
</tr>
<tr>
<td>Honduras</td>
<td>13</td>
<td>44</td>
<td>970</td>
<td>133</td>
<td>66</td>
<td>14,7</td>
<td>15,4</td>
</tr>
<tr>
<td>Mexico</td>
<td>8</td>
<td>27</td>
<td>1184</td>
<td>695</td>
<td>209</td>
<td>5,7</td>
<td>6,1</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>6</td>
<td>39</td>
<td>1290</td>
<td>82</td>
<td>37</td>
<td>34,9</td>
<td>35,4</td>
</tr>
<tr>
<td>Panama</td>
<td>7</td>
<td>19</td>
<td>1317</td>
<td>459</td>
<td>294</td>
<td>4,5</td>
<td>4,8</td>
</tr>
<tr>
<td>Peru</td>
<td>10</td>
<td>72</td>
<td>1121</td>
<td>288</td>
<td>172</td>
<td>6,5</td>
<td>7,4</td>
</tr>
<tr>
<td>Uruguay</td>
<td>10</td>
<td>43</td>
<td>2820</td>
<td>532</td>
<td>272</td>
<td>10,4</td>
<td>11,0</td>
</tr>
<tr>
<td>Venezuela</td>
<td>16</td>
<td>141</td>
<td>1859</td>
<td>610</td>
<td>205</td>
<td>9,1</td>
<td>10,6</td>
</tr>
<tr>
<td>Average LA</td>
<td>12</td>
<td>58</td>
<td>1238</td>
<td>380</td>
<td>144</td>
<td>11,1</td>
<td>11,9</td>
</tr>
<tr>
<td>United States</td>
<td>6</td>
<td>6</td>
<td>368</td>
<td>3837</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**. Number of procedures, duration and cost to start a firm in Latin America (columns 1-3). Official and informal per capita income figures (columns 4-5). Ratio between the cost to start a firm and informal monthly GNP (column 6) and ratio between a total cost measure, incorporating time costs, and informal monthly GNP (column 7). Sources are Schneider (2002) and World Bank (2009A, 2009B).

To finish this section, three examples on capital stock levels and profits from small (typically informal) firms, are given.

In a representative sample of 3700 firms with 5 or less employees in Mexico, McKenzie and Woodruff (2006) report that the median capital stock replacement value across industries is USD 963. In Mexico, typical capital stocks are thus worth less than the costs to go through the registration procedure, from table 2. In the same Mexican data set, the average reported monthly earnings for firms with less than the median capital stock is USD 172 (Woodruff, 2006). In another study from Sri Lanka, De Mel, McKenzie and Woodruff (2008) report that the median level of invested capital for 408 firms is around USD 18013.

13In the latter study, firms with less than 1000 USD in capital stock were targeted, this caused around 6% of the originally selected sample of entrepreneurs to be dropped. The
As an example of informal firm profits, the Brazilian study of informal 1-5 person firms cited in the introduction reports that roughly 75% of firms say they make profits. The average monthly profit of these profit-making firms was USD 314. The profit for firms with remunerated employees was USD 825. For own account firms/workers, that may or may not have non-remunerated employees, profits were USD 235 (SEBRAE, 2005). These entrepreneurial activities are often the main or sole activity of the persons involved, indicating a small room for anything else than consumption expenses.

4 The model

In this section a dynamic model of firm investment, growth and possible formalization is introduced and solved. The firm starts out as informal and the question is if, when and at what firm size the firm will become formal. The modeling is inspired by the framework in Harstad and Svensson (2009).

The production function is simple: production is linear in the capital stock \((k_t)\). As informal the firm produces \(A^i k_t\), if it has formalized, production is instead \(A^f k_t\), where \(A^f > A^i\). It is thus assumed that formality is desirable for the firm. I first solve a dynamic profit maximization problem in sections 4.1 to 4.5. Because the focus throughout section 4 is on one individual firm, heterogeneity between firms is also not introduced until section 4.6, after which I also discuss a possible microfoundation for \(A^i\).

The firm can grow by investing \((i_t)\) in its capital stock. The cost of investing is convex in the size of the investment, \(z^2 i_t^2 / 2\). This gives a profit flow \((\pi_t)\), in the case of the firm being informal, as follows:

\[
\pi_t = A^i k_t - z^2 i_t^2 / 2
\]

The capital stock depreciates at a rate \(\delta\). The growth of the capital stock is therefore

\[
\dot{k}_t = i_t - \delta k_t.
\]

In order to get access to the higher productivity \(A^f\), the firm must pay a formalization fee \(F\) at some time \(T\). After formalization, flow profits equal \(A^f k_t - z^2 i_t^2 / 2\). The firm discounts future profits at a rate \(\rho\).

authors argue that "we believe the resulting sample is representative of a substantial majority of the own-account workers in Sri Lanka" (page 1335, footnote 4).

\(^{14}\)An average exchange rate of 2.86 Reais/USD in October 2003 was used to calculate these numbers.
The basic dynamic problem, in an environment where there are no restrictions on how the firm can finance investment and formalization costs from its own lifetime revenue, is stated below. The effect from different credit restrictions on the problem set-up are discussed in section 6. This discussion is postponed because credit constraints turn out to affect the dynamic analysis in a way which can be handled within the main framework.

4.1 The firm profit maximization problem
An informal firm, starting with an initial capital stock of $k_0$, chooses an investment path, whether it should become formal and the time of formalization ($T$). The firm’s profit maximization problem can be written as:

Choose $i_t, T$ to Max $\int_0^T \left( A^f k_t - \frac{z}{2} \frac{1}{t^2} \right) e^{-\rho t} dt + \int_T^\infty \left( A^f k_t - \frac{z}{2} \frac{1}{t^2} \right) e^{-\rho t} dt - F e^{-\rho T}$

s.t. $\dot{k}_t = i_t - \delta k_t$ and $k(0) = k_0$ (3)

The problem can be solved in two steps. First, we use the principle of optimality to solve backwards for the formal- and then the informal investment path (assuming that $T$ exists). We also derive the investment path if $T$ does not exist. By using the investment path assuming that formalization does take place, we then determine when the firm wants to formalize by solving for the optimal $T$. If such a $T$ exists we then know the optimal capital accumulation path. If it does not exist, the firm is informal forever.

4.2 Optimal investments
Assume that $T$ exists. Solving backwards, the "formal problem" takes the capital stock at time $T$, defined as $k_T$, as an initial condition, and is solved for the investment path from $T$ to $\infty$. We get a formal investment function $i_{\text{formal}}$ and a continuation value $V_{\text{formal}}$, which is the optimal profit from $T$ and onwards. $V_{\text{formal}}$ will be a function of both $T$ and $k_T$. The profit maximization problem is\textsuperscript{15}:

Choose $i_t$ to Max $\int_T^\infty \left( A^f k_t - \frac{z}{2} \frac{1}{t^2} \right) e^{-\rho t} dt$

s.t. $\dot{k}_t = i_t - \delta k_t$ and $k(T) = k_T$

By defining the present-value Hamiltonian $H = \left( A^f k_t - \frac{z}{2} \frac{1}{t^2} \right) e^{-\rho t} + \lambda_t (i_t - \delta k_t)$, where $\lambda_t$ is the present value Lagrange multiplier on the capital accumulation

\textsuperscript{15}$k_T$ is not a choice variable in the overall problem, it is only introduced as an auxiliary variable when we solve the formal and informal problems separately.
constraint, and applying the first order conditions \( \frac{\partial H}{\partial i} = 0, \frac{\partial H}{\partial k} = -\frac{\partial \lambda}{\partial t} \) and

the transversality condition \( \lim_{t \to \infty} (\lambda_t k_t) = 0 \), we get the optimal solution:

\[
\begin{align*}
\dot{i}_{\text{formal}} &= \frac{A_f}{z (\delta + \rho)} \\
\dot{k}_{\text{formal}} &= \dot{k}_T e^{-\delta (t-T)} + \frac{A_f}{z \delta (\delta + \rho)} (1 - e^{-\delta (t-T)}) \\
V_{\text{formal}} (T, \dot{k}_T) &= e^{-\rho T} \left( \frac{A_f \dot{k}_T}{\delta + \rho} + \frac{(A_f)^2}{2 \rho (\delta + \rho)^2} \right) \\
k_{\text{formal}}(\infty) &= \frac{A_f}{z \delta (\delta + \rho)} \\
n(4)
\end{align*}
\]

The firm invests a constant amount each "period". The capital stock converges to its steady state value of \( k_{\text{formal}}(\infty) = \frac{A_f}{z \delta (\delta + \rho)} \), at which depreciation and investment offset each other. The constant investment rate is due to the convexity of investment costs - the firm wants to spread investment over time. The investment rate increases in the productivity parameter \( A_f \) and decreases in the cost of investment \( z \), the depreciation rate of capital \( \delta \) and the rate of time preference \( \rho \).

The informal investment path, for a given \( T \), can in turn be determined by solving for the investment path that takes the firm from \( k_0 \) to \( \tilde{k}_T \) and then maximize total profits with respect to \( \tilde{k}_T \):

Choose \( i_t \) and \( \tilde{k}_T \) to Max

\[
\int_0^T \left( A^i_i k_t - \frac{z}{2} i_t^2 \right) e^{-\rho t} dt + e^{-\rho T} V_{\text{formal}} (T, \tilde{k}_T)
\]

s.t. \( \dot{k}_t = i_t - \delta k_t \), \( k(0) = k_0 \) and \( k(T) = \tilde{k}_T \).  

The investment path is derived as above, the only difference being the terminal constraint on capital (instead of a transversality condition). Having solved for the optimal informal \( i_t \) - and \( k_t \)-paths as functions of \( \tilde{k}_T \), and plugged these back into the profit function, we integrate to get the optimal value of informal profits as a function of \( \tilde{k}_T \).  The optimality condition with respect to \( \tilde{k}_T \) is that the loss in informal profits from increasing \( \tilde{k}_T \) should be exactly offset by a gain in formal profits:

\[
\frac{d}{d\tilde{k}_T} \left( \int_0^T \left( A^i_i k_t (\tilde{k}_T) - \frac{z}{2} (i_t (\tilde{k}_T))^2 \right) e^{-\rho t} dt + e^{-\rho T} V_{\text{formal}} (T, \tilde{k}_T) \right) = 0 \quad (6)
\]

\(^{16}\)In solving the problem, a non-explosive path of investment is profit-maximizing. Other investment paths, that fulfill the differential equations for \( i_t \) and \( k_t \) stemming from the first order conditions on the Hamiltonian, can be ruled out for optimality reasons (and hence do not fulfill \( \lim_{t \to \infty} (\lambda t k_t) = 0 \)).

12
The optimal $k_T$ is plugged back into the solution for $i_t$ and $k_t$, which after simplification become

$$i_{t}^{\text{formalization}} = \frac{A^i}{z(\delta + \rho)} + \frac{A_f - A_t^i}{z(\delta + \rho)} e^{(\delta + \rho)(t - T)}$$

$$k_{t}^{\text{formalization}} = k_0 e^{-\delta t} + \frac{A^i (1 - e^{-\delta t})}{z \delta (\delta + \rho)} + \frac{(A_f - A_t^i) \left( e^{(\delta + \rho)(t - T)} - e^{-(\delta + \rho)T - \delta t} \right)}{z (\delta + \rho) (2\delta + \rho)} \tag{7}$$

This investment path starts out close to $A^i / (z(\delta + \rho))$, and then increases up to the level of formal investments at $T$, i.e. $A_f / (z(\delta + \rho))$. Investment increases close to formalization because the marginal value of capital is high after formalization, this makes the firm willing to decrease its profits by accumulating more capital, while still being informal.

Now assume that $T$ does not exist. The firm is then informal forever. Solving this problem is identical to solving the formality problem above, but productivity is $A^i$, time runs from 0 and the initial capital stock is $k_0$. The "ever-informal" problem is:

Choose $i_t$ to Max \( \int_0^\infty \left( A^i k_t - \frac{z}{2} i_t^2 \right) e^{-\rho t} dt \)

s.t. $\dot{k}_t = i_t - \delta k_t$ and $k(0) = k_0$

The solution, obtained as in the formal problem above, is:

$$i_{t}^{\text{informal}} = \frac{A^i}{z(\delta + \rho)}$$

$$k_{t}^{\text{informal}} = k_0 e^{-\delta t} + \frac{A^i}{z \delta (\delta + \rho)} (1 - e^{-\delta t}) \quad k_{t}^{\text{informal}} = \frac{A^i}{z \delta (\delta + \rho)} \tag{8}$$

As for the investment path once formal, the investment rate is constant and the capital stock converges to a steady state value, $k_{\infty}^{\text{informal}} = \frac{A^i}{z \delta (\delta + \rho)}$. This capital stock is lower than if the firm had been formal, because productivity is lower.

### 4.3 Solving for the formalization time $T$

If $T$ exists, the investment path before and after formalization is given above (expressions 7 and 4, respectively. The optimal $T$ can be derived by recognizing that, at the time of formalization, it must be that formalization is just as attractive as remaining informal. This determines the capital stock at which the firm wants to formalize, which in turn, with the capital accumulation prior to
formalization $k_t^{\text{formalization}}$ given in (7), determines $T$. We get that formalization takes place when

$$
\frac{d}{dT} \left( \int_0^T \left( A^i k_t - z i^2 t \right) e^{-\rho t} dt + \int_T^\infty \left( A^f k_t - z i^2 t \right) e^{-\rho t} dt - F e^{-\rho T} \right) = 0 \quad (9)
$$

As discussed above, the pre-formalization investment rate approaches the formal investment rate as $t \to T$. At $T$ these effects cancel out and the condition in (9) simplifies to $A^i k_T - A^f k_T + \rho F = 0$. The optimal capital stock at formalization, defined as $k^F$, becomes

$$
k^F = \frac{\rho F}{A^f - A^i} \quad (10)
$$

We get $T$ by equating the optimal capital accumulation path at $t = T$, i.e. $k_T^{\text{formalization}}$, with $k^F$:

$$
k_0 e^{-\delta T} \frac{A^i (1 - e^{-\delta T})}{z (\delta + \rho)} + \frac{(A^f - A^i) (1 - e^{-(2\delta + \rho)T})}{z (\delta + \rho) (2\delta + \rho)} = \frac{\rho F}{A^f - A^i} \quad (11)
$$

This equation implicitly defines the optimal time of formalization $T$. Formalization means a promise of future higher profits. The firm that formalizes builds a higher capital stock while informal, in anticipation of such profits. Because $i_t^{\text{formalization}} > i_t^{\text{informal}}$, this period is thus associated with losses compared to the "ever-informal" path. There is a certain amount of losses/extra investment that can be sustained in anticipation of formalization, this gets reflected in the amount of capital that is optimally accumulated prior to formalization, i.e. the LHS in (11).

The formalization decision also depends on at what capital stock paying $F$ is optimal. The first order condition in (9) implies that the marginal gain from formalization, which is $(A^f - A^i)$ times the capital stock, should equal the marginal loss of not delaying formalization, i.e. $\rho F$.

It should be observed at this stage that although we have not restricted the time of payment of $F$ in any sense, the firm does not want to pay the formalization fee at once. This is because it is only beneficial to pay $F$ once a certain capital stock/firm size has been reached, and getting there is costly due to the convexity of investment costs.

4.3.1 Existence of $T$

Determining under what conditions $T$ exists completes the solution to the dynamic problem. Proposition 1 below states the full conditions for when a firm formalizes. The main idea in deriving this proposition is to let $T \to \infty$ in
the LHS of expression (11), this gives an auxiliary maximum level of capital
\[k_{\infty}^{\text{formalization}} = \frac{A^i(\delta + \rho) + \delta A^f}{z\delta (\delta + \rho) (2\delta + \rho)}\] in anticipation of formalization, then compare this capital level with the RHS in (11). Appendix 1 gives further details.

**Proposition 1**: A firm that starts with a capital level \(k_0\) less than \(k_{\infty}^{\text{informal}} = \frac{A^i}{z\delta (\delta + \rho)}\) will become formal if and only if the formalization cost \(F\) is less than or equal to \(\bar{F} = \frac{(A^f - A^i) (A^i (\delta + \rho) + \delta A^f)}{z\delta \rho (\delta + \rho) (2\delta + \rho)}\). This threshold is increasing in \(A^f\), decreasing in \(z\), \(\delta\) and \(\rho\) and increasing in \(A^i\), then decreasing. For firms that start with \(k_0\) larger than \(k_{\infty}^{\text{informal}}\), formalization will take place if and only if \(F \leq \bar{F} + G(k_0)\), where \(G(k_0)\) is positive and a strictly increasing function of \(k_0\).

The next subsection states the full solution. A second proposition is then presented, after which the basic comparative statics and intuition of firm formalization are discussed. The discussion of the second part of proposition 1, the \(k_0\)-dependence, is postponed until section 5.

### 4.4 The full solution to the dynamic problem

The solution to the dynamic problem can be stated as follows: If the conditions in proposition 1 are satisfied, there exists a formalization time \(T\) which is the solution to equation 11. In this case the firm follows the formalization investment path \(i_{\text{formalization}}(t)\) from expression 7), then switches at time \(T\) to the formal investment path \(i_{\text{formal}}(t)\) from expression 4). Such an investment path is shown in figure 1. If instead proposition 1 would not have been satisfied, the firm would have followed an "informal-ever" investment path \(i_{\text{informal}}(t)\) from 8, the broken line in figure 1).
4.5 Comparative statics of the dynamic problem

Proposition 1 was derived from expression (11). An alternative approach to the above is to use expression (11) to analyze the comparative statics of the time of formalization $T$. The same parameter changes that make formalization "easier" (reflected in an increase in $F$) also imply a smaller $T$.

**Proposition 2**: The formalization time $T$ is a function of all the parameters of the problem: $T(F, z, \delta, \rho, A^f, k_0, A^i)$. It is increasing in $F$, $z$, $\delta$, and $\rho$ and decreasing in $A^f$ and the initial capital stock $k_0$. It is decreasing in $A^i$ for small values of $A^i$, then increasing.

Increases in the formalization fee $F$ will make the necessary capital accumulation take longer time. An increased cost of investing $z$ slows down the growth of the capital stock. Preformalization investments also decrease unambiguously in the depreciation rate $\delta$ and in the discount rate $\rho$. In addition, an increase in $\rho$ makes firms want to postpone formalization (the RHS in 11 increases), which makes $T$ increase further. An increase in $A^f$ strengthens the incentive to invest (LHS of 11). In addition, it decreases the level of capital $k^F$ at which formalization becomes advantageous (RHS of 11). Both effects speed up formalization. The initial capital stock adds to the capital stock obtained by investing, and $T$ is therefore smaller the higher is $k_0$.

With respect to the informal productivity level $A^i$ there are two effects: an *investment effect* and a *threshold effect*. An increase in $A^i$ means more investment and capital accumulation (LHS in 11) but also that formalization
becomes less advantageous (RHS in 11). For small values of $A^i$ (in comparison to $A^f$) the investment effect dominates and formalization becomes easier ($F - F'$ increases, proposition 1) and faster ($T$ decreases, proposition 2). For large values of $A^i$ the threshold effect instead dominates.

The response in $T$ to changes in the parameter values implies that there are two effects on the investment path when a parameter changes. Consider an increase in $A^f$. This produces a direct effect by which $t_{\text{formalization}}^f$ in (7) increases, for a given $T$. In addition, there is an indirect effect through a smaller formalization time $T$, which further increases investment at any moment in time. In figure 1, these direct and indirect effect could be depicted as a formalization (pre-$T$) investment path at a higher level and with a higher slope at each point in time, a shift to the left in $T$, and a shift up in the formal (post-$T$) investment level.

This unambiguous multiplicative effect is also present (but goes in the other direction) for changes in $z$, $\delta$, and $\rho$.

Before analyzing these results further, heterogeneity between firms is introduced in section 4.6 and a microfoundation for the informal productivity parameter $A^i$, connected to a penalty policy vis-a-vis informal firms, is provided in section 4.7.

### 4.6 Introducing firm heterogeneity

The discussion so far has concerned one firm. In order to allow a discussion in section 5 about firms that become formal versus those that do not, an assumption about firm heterogeneity is introduced. Specifically, let firms be indexed by $j$ and assume that there is a firm-specific "ability" parameter $\theta_j$ that multiplies two baseline productivity parameters $A^I$ and $A^F$. For the sake of simplicity, let $\theta_j$ be uniformly distributed on the unit interval, $0 \leq \theta_j \leq 1$. The baseline parameters $A^I$ and $A^F$ can be interpreted as the maximum productivities of the informal and formal sectors, respectively. A number of reasons why these may differ was outlined in the introduction. Each firm, through its ability parameter, then has its own productivity in relation to $A^I$ and $A^F$: as formal it is $A^f_j = \theta_j A^F$. The productivity of the same firm $j$ while informal, $A^i_j$, contains an additional component, discussed in the next section.

### 4.7 The informal productivity $A^i_j$

As discussed in the introduction, there are potentially many different reasons for productivity differences between informality and formality. One often discussed informality/formality difference is the risk of being caught for operating
"illegally" in the informal sector. This section will formally model how this can affect firm productivity. The aim is both to provide one possible microfoundation of firm investment behavior in the informal sector, and to provide a basis for a discussion of policy vis-a-vis informal firms.

The risk of being detected and penalized by authorities for operating "illegally" is an often-used characterization of the informal firm environment, both by informal entrepreneurs themselves (Tokman, 1992) as well as in economic models (Loayza, 1995). One reason for such action by the authorities is that informal firms do not pay taxes, in order to increase tax revenue the government therefore wants firms to formalize. Another rationale is that formal firms pressure the authorities to deal with informality, claiming that competition from non-compliers is "unfair".

An explicit story for how such a policy towards informal firms affects the informal productivity $A_j$ is through the time use of informal entrepreneurs (see section 2 for a discussion). Let the productivity of the informal firm, if it can operate without hiding, be $\theta_j A^I$ from above. Let $l$ be the fraction of the informal entrepreneur’s unitary time endowment spent trying to avoid detection, rather than in production, let $p(l)$ be the resulting probability of not getting caught and let $x$ be the fraction of output which is taken from the informal entrepreneur if caught (the penalty/policy parameter). The expected productivity when operating informally becomes $p(l)\theta_j A^I (1 - l) + (1 - p(l)) \theta_j A^I (1 - l) (1 - x)$, which can be rewritten as

$$\theta_j A^I (1 - l) - x \theta_j A^I (1 - l) (1 - p(l))$$

The first term reflects production, the second term the effect from penalties. Let the probability of not being detected be $p(l) = \sqrt{l}$. This function fulfills the natural requirements that $p(0) = 0$, $p(1) = 1$ and also $dp/dl_{l=0} = \infty$ and $d^2p/dl^2 < 0$. By solving for the optimal time allocation and detection probability from the first order condition $-\theta_j A^I + p(l)x \theta_j A^I (1 - l) + x \theta_j A^I (1 - p(l)) = 0$, we get the informal productivity parameter\footnote{The capital stock $k$ is omitted because the time allocation decision to maximize expected "per period" production is static and independent of the dynamic investment decision in (3).}. It is a strictly decreasing and convex function $h(x)$ of the penalty parameter $x$, where $h(0) = 1$, multiplied with $\theta_j A^I$:

$$A_j = \theta_j A^I h(x)$$

The penalty parameter thus affects productivity negatively. Firms can shed themselves from the worst case, by allocating time on "hiding" rather than on production ($h(x)$ is always larger than $1 - x$)\footnote{We could use a more general function for the probability: $p(l) = l^\xi$ with $0 < \xi < 1$. The parameter $\xi$ would in a sense reflect the strength of enforcement of penalties $x$. A lower value of $\xi$ would mean that even small amounts of time used for "hiding" are very effective in avoiding detection and could be interpreted as weak enforcement.}. Figure 2 shows the resulting

\begin{equation}
\theta_j A^I (1 - l) - x \theta_j A^I (1 - l) (1 - p(l))
\end{equation}
informal productivity, $A_i^j = \theta_j A^I h(x)$ as a function of $x$ for an individual firm $j$. The formal productivity $A_f^j = \theta_j A^F$ is also shown.

The effects from penalties on formalization are discussed in section 5.

\[ A_i^j = \theta_j A^I h(x) \]

Figure 2. The informal and formal productivities of an individual firm $j$.

5 Analysis of the model

What does the dynamic model imply in terms of firm formalization and investment? This section discusses a few predictions, starting out with a proposition about which firms that formalize, the investment paths, time of formalization and firm size at formalization.

5.1 Characteristics of firms that become formal

Expression (11) is repeated for convenience, disregarding the effect from initial capital$^{20}$:

\[
\frac{A_i^j \left(1 - e^{-\delta T}\right)}{z\delta (\delta + \rho)} + \frac{\left(A_f^j - A_i^j\right) \left(1 - e^{-(2\delta + \rho)T}\right)}{z (\delta + \rho) (2\delta + \rho)} = \frac{\rho F}{A_f^j - A_i^j}
\]  

(11')

As derived in sections 4.6-4.7, the productivities are $A_i^j = \theta_j A^I h(x)$ and $A_f^j = \theta_j A^F$.

$^{20}$If not explicitly stated, I assume that the initial capital of firms is small, such that there is no $k_0$-dependence in whether firms formalize or not (see proposition 1).
Proposition 3. Firms with an ability parameter $\theta_j$ above a threshold value $\theta_{\text{formalization}}$, i.e. firms in the range $\theta_{\text{formalization}} \leq \theta_j < 1$, become formal. For such firms, the larger is $\theta_j$ the larger is investment, the faster is formalization and the smaller is the firm size at which formalization takes place.

The intuition for this proposition is straightforward: firms with high $\theta_j$ both invest more due to a higher productivity (LHS of 11 increases), and they have more to gain more from formalization (RHS of 11 decreases). In a cross section of firms, we thus should not only observe that it is high ability/productivity firms that become formal, but furthermore that their firm size at formalization is smaller and the time from firm start-up to formalization is shorter.

The threshold value $\theta_{\text{formalization}}$ is derived by plugging in the full expressions for firm productivities ($A_j^f = \theta_j A^f h(x)$ and $A_j^i = \theta_j A^i$) in the formalization criterion derived in the first part of proposition 1 and solving for $\theta_j$, which gives

$$\theta_{\text{formalization}} = \sqrt{\frac{A^F z \delta (\delta + \rho)(2\delta + \rho)}{(A^F - A^f h(x))(A^i h(x)(\delta + \rho) + \delta A^F)}}$$

(14)

5.2 Penalties

5.2.1 Policy maker

Before analyzing the effects from penalties on formalization, a highly relevant question is: Who is this policy maker? So far, the penalty policy has been connected to a somewhat diffuse "authority".

One interpretation of the penalty parameter $x$ is that it is the government that sets (and enforces) such penalties. It is then assumed that the government can audit informal firms and penalize them for operating illegally. In practice, this could take place through "benevolent" tax officers, police, local authorities etc. One reason for such audits to take place may be that the government wants to increase tax revenue by making firms formal or that there is some negative externality from informal production.

An alternative view on policy, much different, is when there is no government in the traditional sense. Indeed, we are studying the informal sector which, by definition, consists of unregistered firms. The penalty parameter $x$ perhaps instead is collected by "malevolent" police, corrupt bureaucrats, local mafias etc. (De Soto, 1989; Tokman, 1992). The likely aim then is not to speed up formalization, but to maximize bribe revenue from informal firms. Although I do not provide a formal analysis of such a case, there is no reason to believe that penalties would be set as in the "benevolent" case. Instead, one can hypothesize about the effects from short time horizons of "collectors" ($\Rightarrow x \uparrow$), lack of commitment to refrain from collecting more bribes ($\Rightarrow x \downarrow$), no desire that firms
should become formal and disappear from the "tax base" ($\Rightarrow x \uparrow$), risk of being detected if collecting too much ($\Rightarrow x \downarrow$), and so on.

In the following section, the policy maker is the government and the optimal penalty to maximize firm formalization is derived.

5.2.2 Effects from penalties on formalization

By analyzing $\theta_{\text{formalization}}$, we can study how the government policy parameter vis-a-vis informal firms affects formalization. Whereas $\theta_{\text{formalization}}$ increases in $F$, $z$, $\delta$ and $\rho$ and decreases in $A^F$, there is an ambiguous effect with respect to $A^I h(x)$. This effect was observed in analyzing proposition 2 and is restated here:

**Proposition 4.** The effect from the penalty parameter on the threshold for formalization $\theta_{\text{formalization}}$ is U-shaped, first decreasing in $x$ for small values of $x$, then increasing. The penalty parameter that minimizes $\theta_{\text{formalization}}$, i.e. that maximizes the amount of firms that formalize, is $x = 0$ when $A^I \leq \frac{A^F \rho}{2 \delta + \rho}$, in an intermediate range of $A^I$ it is given by the $x$ that solves $h(x) = \frac{A^F \rho}{2 A^I \delta + \rho}$ and it is $x = 1$ when $A^I \geq \frac{3 \sqrt{3} A^F \rho}{4 \delta + \rho}$.

The policy maker can affect the incentive to formalize through the penalty on informal production. For small values of $x$, the threshold effect will dominate - formalization becomes more attractive when penalties are increased. This is seen in (11'), where the level of capital at which the firm optimally formalizes (the RHS) goes down. For large penalties it is instead the case that the investment effect will dominate - firms will accumulate less capital and therefore not be able to become formal. This is the LHS of 11'.

The penalty that maximizes the amount of firms that formalize is derived through the necessary condition for a minimum on $\theta_{\text{formalization}}$, i.e. $\partial \theta_{\text{formalization}} / \partial x = 0$. This condition gives $h(x) = \frac{A^F \rho}{2 A^I \delta + \rho}$, which in turn must lie between $h(0) = 1$ and $h(1) = \frac{2}{3 \sqrt{3}}$. This gives the proposition.

A restatement of proposition 4 is that a policy designed by a government to incentivize firms to become formal, should be conducted with a "carrot and stick" approach: neither too mild nor too tough. The accumulation- and threshold effects will be balanced and the amount of firms that become formal is maximized.

As an illustration, assume that $A^I$ equals $\frac{1}{2} A^F$ and that $\delta = \rho$. Because $\frac{1}{4} < A^I < \frac{3 \sqrt{3}}{8}$ we have an interior solution and we get $h(x) = \frac{1}{2}$. Using the
\( h(x) \)-function (appendix 2), we get that \( x^+ \approx 0.68 \) maximizes the amount of formalization. The graph below on the shape of \( \theta_{\text{formalization}} \) illustrates that lower penalties will result in much less formalization, whereas higher penalties do not affect the degree of formalization as much.

![Graph showing the effects of penalties on formalization](image)

**Figure 3.** An illustration of the effects from penalties on the minimum ability threshold for formalization, \( \theta_{\text{formalization}} \).

### 5.3 The formal sector productivity \( A^F \)

Although we have not made explicit the formal sector productivity parameter \( A^F \), nor specified it as a policy parameter, it is worth pointing out that increases in \( A^F \) has two effects. First, investments increase (investment effect, LHS of 11). Second, the firm size at which formalization becomes beneficial goes down (threshold effect, RHS of 11). Obviously, decreasing \( A^F \) has the opposite effect.

The models by Loayza, 1995 and Garcia Penalosa and Turnovsky, 2005 include taxes and public goods as determinants of the formal sector productivity. Higher taxes and less efficient public goods provisioning in the formal sector both act to increase informality. In the present paper, we can think of these policy parameters as potential determinants of \( A^F \). The preceding paragraph clarifies two channels through which investment in the informal sector and formalization is discouraged by higher taxation and less efficient public goods provisioning in the formal sector.

### 5.4 The aggregate of firms

What does the long run firm size distribution predicted by the model look like? Does it resemble actual firm size distributions in developing economies?
Over the long run, informal firms converge to a (firm specific) size
\[ k_{\text{informal}} = \frac{\theta_j A^I h(x)}{z \delta (\delta + \rho)} \]
and formal firms to
\[ k_{\text{formal}} = \frac{\theta_j A^F}{z \delta (\delta + \rho)} \].

The firm with an ability parameter marginally lower than \( \theta_{\text{formalization}} \) thus reaches a much smaller size than had the ability parameter been somewhat larger. The size of the firm size gap, i.e.
\[ \frac{\theta_{\text{formalization}} A^F}{z \delta (\delta + \rho)} - \frac{\theta_{\text{formalization}} A^I h(x)}{z \delta (\delta + \rho)} \],
is increasing in \( F, \rho, A^F \) and decreasing in \( A^I h(x), z \) and \( \delta \). Together with the fact that \( \theta_{\text{formalization}} \) increases in \( F \) and that \( h(x) \) decreases in \( x \), we can state the following proposition:

**Proposition 5.** In the long run, the model displays a low-end range, 0 \( \leq \theta_j \leq \theta_{\text{formalization}} \) of small informal firms and a high-end range, \( \theta_{\text{formalization}} \leq \theta_j \leq 1 \), of large formal firms. There is a "missing middle" in firm sizes, and the size of the gap is increasing in formalization costs \( F \) and in penalties \( x \).

Tybout (2000) documents firm size distributions for a number of developing economies and finds evidence of a "dual structure", with a large proportion of very small firms, a "missing middle" and then a few large firms. This contrasts with typical high-income countries. The author further argues that "small producers frequently operate partly or wholly outside the realm of government regulation" (Tybout 2000, page 15), discussing costs of dealing with the government as one explanation for the observed pattern. The present model shows how profit maximization behavior of firms, with large costs of entry into formality, can generate such a "missing middle". It also delivers predictions about the size of this firm size gap. Figure 4 shows the long run firm sizes as predicted by the model.

![Figure 4. Long term firm size distribution.](image-url)
5.5 The dependence on initial capital

Heterogeneity between firms was introduced along an ability/productivity dimension, resulting in predictions about which firms that formalize. The focus of the poverty trap literature is instead to study how the initial wealth distributions matters for the future wealth distribution. Translated to the present paper, the question is if differences in initial capital $k_0$ can explain differences in long-run firm sizes. At face value, the answer to this question should be no: there are no explicit credit constraints in the model, therefore equally productive firms should converge to the same steady state in the long run (Banerjee, 2001; McKenzie and Woodruff, 2006).

Proposition 1 states that initial capital plays a role however. I disregard the ability differences for now and assume all firms are equal except for differences in $k_0$. Whenever $k_0$ is larger than $k_{\text{informal}}$, the maximum level of capital that can optimally be accumulated is larger than $k_{\text{formalization}} = \frac{A^i(\delta + \rho) + \delta A^f}{z\delta(\delta + \rho)(2\delta + \rho)}$. The maximum level of capital is also increasing in $k_0$. These statements were proven in appendix 1.

A high initial capital stock gives a firm an "initial cost advantage" in reaching a certain size. That is, it is (initially) less costly for a firm starting with a high $k_0$ to reach a certain (larger) capital level than if the firm had started with a low $k_0$. This effect vanishes over time because the initial capital depreciates at a rate $\delta$. The way in which a firm starting with high $k_0$ takes advantage of the effect is to invest heavily in the beginning and then formalize early (at capital level $k^F$ from $10$). Early formalization makes the extra investment worthwhile because formalization profits come closer in time. A firm starting with a lower $k_0$ has no possibility to take advantage of the effect from initial capital because to reach $k^F$ would mean even higher initial investment losses and/or later formalization, implying a lower net present value of formality.

An alternative explanation for the fact that the cost advantage due to a high $k_0$ is only initial is as follows: A firm that is informal forever converges to a capital stock $k_{\infty}^{\text{informal}}$ at which investments just compensate depreciated capital. For a firm that formalizes at some time $T$ far in the future, it is costly to deviate much from $k_{\infty}^{\text{informal}}$ prior to formalization. Therefore, for a $T$ that is large, investments will mimic the "ever-informal" case where depreciated capital is just replaced, as long as $T$ is not close in time. This in turn means that the effect from the initial capital stock will vanish prior to $T$ and, as a result, not affect the decision whether to formalize or not.

The result that $k_0$ matters resembles the poverty trap literature. A certain formalization level $k^F$ that would be prohibitive for firms with $k_0$ smaller than $k_{\infty}^{\text{informal}}$, will still allow for a range of firms with high enough initial capital to formalize and converge to $k_{\infty}^{\text{formal}}$. 

24
6 Extensions

This section discusses some modifications to the model in section 4. I abstract from firm heterogeneity in the presentation below, to save on notation.

6.1 The effect of borrowing and savings constraints

So far, nothing hinders the firm from making losses while informal. These losses can come from three sources: early in the process due to investment costs exceeding production revenue, due to higher investment costs in anticipation of formalization, as well as in the instant when $F$ is paid. In the formulation in (3), the only concern is that the net present value of revenue exceeds the net present value of costs\(^{21}\).

It is beyond the scope of this paper to review the literature on credit\(^{22}\). Instead, the introduction of a no-borrowing constraint in this model is, somewhat innocently, motivated by the observation that most informal firms do not use credit at all. In the Brazilian representative sample of small informal firms, 6\% of firms had used credit during the 3 months prior to the study and 17\% of firms had any debt outstanding whatsoever (SEBRAE, 2005). These numbers seem to be quite typical. In a study of six African countries, 2\% of the micro firms (1-5 employees) and 7\% of small firms (6-25 employees) had received a loan in the year prior to the study. 16\% of each group had debt in the informal finance sector (Bigsten et al., 2003)\(^{23}\). Levy (1993) reports small percentages of Tanzanian and Sri Lankan micro firms (1-5 employees) with formal credit.

In two recent empirical papers estimating returns to capital, McKenzie and Woodruff (2006) and de Mel, McKenzie and Woodruff (2008) find evidence on binding credit constraints for small informal firms in Mexico and Sri Lanka.

In addition to borrowing constraints, it has also been argued that savings constraints may be important in developing countries (Duflo and Banerjee, 2007; Dupas and Robinson, 2009). In the model presented here, this would mean that informal firms cannot save from current profits to pay the formalization fee. The impact of such a constraint is discussed below.

---

\(^{21}\)One can think of the firm as moving profits between time periods at an internal interest rate of $\rho$. Making a loss of a certain amount at time $t$, must be compensated by a profit which is a factor $e^{\rho \Delta t} \approx 1 + \rho \Delta t$ higher at time $t + \Delta t$.

\(^{22}\)Banerjee, 2001, reviews a number of "stylized facts" about credit markets in developing countries.

\(^{23}\)As the authors point out, not having credit does not mean that firms do not want it and 64\% of the micro firms had either been denied credit or say they would be denied if they apply. It should also be pointed out that several studies report trade credit to be an important source of financing for small firms.
6.1.1 No borrowing to finance investment

First assume that the firm cannot borrow to finance investment. Because the firm would like to invest at least an amount $i_{\text{informal}}$ from the start, the no-borrowing constraint will clearly be binding for firms that start with a small initial capital stock (imagine $k_0$ close to zero, this gives a profit flow in the unconstrained case of $A^i k_0 - \frac{z}{2} (i_{\text{informal}})^2 < 0$, and hence a need to borrow). Formally, the following constraint should be added to the problem formulation in (3):

$$A^i k_t - \frac{z}{2} i_t^2 \geq 0$$

(15)

Adding this constraint will not affect the formalization decision. Because investment in the capital stock is perfectly divisible and the return on small investment amounts is very high, it will be optimal for the firm to forsake small current profits against future higher profits generated by a larger capital stock. The firm will invest as much as possible, i.e. $i_t = \sqrt{2 A^i k_t / z}$, until the constraint in (15) no longer binds. The firm "bootstraps" its way out of the borrowing constraint. The time it takes depends on $k_0$, but the long run capital stock and formalization decision will not be affected.

6.1.2 Limited borrowing to finance the formalization cost $F$

The introduction of a constraint with respect to the financing of $F$ is more complicated. A shortcut to studying the impact of a borrowing constraint in financing $F$ is as follows: Assume a fraction $\gamma$ of formalization costs can

---

24Formally, the following reasoning leads to the irrelevance of initial borrowing constraints for the formalization decision. First, solving the informal-ever problem with the constraint in (15) gives that the optimal investment path is one in which (15) binds only initially, after which the investment in (8) applies. Second, formalization means an additional incentive to invest so firms will invest as much as they can initially and, after some time $t$, escape the no-borrowing constraint. Third, knowing that (15) will ultimately not apply, the problem from $t$ is the same as the problem solved in (3), only initial conditions differ. As before, it is solved backwards by using the principle of optimality. The resulting optimal investment path from $t$ and onwards will be given by (7), the attainable capital stock as $T \to \infty$ as before and the formalization criterion is given by the first part of proposition 1.

25The model solved in this paper assumes that there is no other asset than the firm's own capital stock in which to invest: no financial saving can be accumulated. Given the evidence on little access to savings devices in the informal sector, this assumption is not implausible. Introducing a no-borrowing constraint to finance $F$ then implies that firms would have to save "in the mattress", i.e. at zero % interest rate, in anticipation of formalization. Introducing a savings control variable ($s_t$), we get that with zero interest the total savings before formalization should at least equal $F$, i.e. $\int_0^T sdt \geq F$. Solving the model with this constraint turns out to be very complicated. It is likely however that the period leading up to formalization would imply a trade-off between investment and "saving in the mattress", implying less capital accumulated and, as result, a smaller possibility for formalization.
be financed as before, by the firm’s own financing. A fraction \((1 - \gamma)\) of the formalization cost can be financed by borrowing, at the instant of formalization, at an interest rate of \(r\). This results in a "per-period" interest payment of \((1 - \gamma) rF\), from \(T\) and onwards. Assuming that the amount borrowed is rolled over indefinitely, the only change to the problem is in the total profit expression in (3), which changes to become

\[
T \int_0^T \left( A^i k_t - \frac{z_i^2}{2} t^2 \right) e^{-\rho t} dt + \int_T^\infty \left( A^f k_t - \frac{z_i^2}{2} t^2 - (1 - \gamma) rF \right) e^{-\rho t} dt - \gamma F e^{-\rho T}
\]

The solution is only affected through condition (9), when solving for the optimal level at which to formalize. The formalization criterion becomes

\[
k^F = \frac{\gamma \rho F + (1 - \gamma) rF}{A^f - A^i} \quad (10')
\]

If the firm borrows at a high interest rate \(r > \rho\), the more difficult will formalization become. The smaller the fraction \(\gamma\) that cannot be financed via own financing, the less likely is formalization.

Expression 10' gives a tractable expression through which to analyze credit constraints. All of the analysis in section 5 remains unchanged, we can think of the credit constraint as being represented by a larger value on the formalization cost \(F\).

The representation above is admittedly a simplified way to introduce a credit constraint. In relation to the literature on poverty traps, one could imagine that the interest rate \(r\) at which the firm can borrow depends on initial capital/initial establishments, such that firms that started off at higher capital stocks represent less of a moral hazard risk to lenders. Such a specification would effectively mean that firms with larger initial capital face lower effective formalization costs, and act as a source of a poverty trap\(^\text{26}\).

### 6.2 An alternative view on penalties

It is often argued that informal firms run a larger risk of detection if they grow, and therefore prefer to stay small. The model can be modified to investigate this argument. Instead of firms spending time on hiding, we now modify the original informal production function. By writing informal production as \(A^i k_t \left(1 - \frac{\beta}{2} k_t\right)\), rather than \(A^i k_t\), we explicitly recognize that as the firm grows, a larger fraction of output will be captured by the authorities (we

\(^{26}\)That is, the combination of the non-convexity \((F)\) and higher interest rates for firms with less initial capital could generate a long-run distribution of capital displaying informal-ever firms with \(k_{\infty}^{\text{informal}}\) and formal firms with \(k_{\infty}^{\text{formal}}\).
think of penalties $x$ incorporated into the $\beta$-term). Although it seems likely that informal firms will now grow (even) less, the incentive to formalize is also stronger than before.

What does this modification to the problem in (3), stated in appendix 3, yield? Although the solution is somewhat complex, parametrizing and comparing it to the model in section 4 gives at hand that for small values of $A^i$ (in comparison to $A^f$), it is the investment effect that dominates. That is, informal firms now simply cannot grow and will formalize to a lesser extent. However, for values of $A^i$ close to $A^f$, the opposite becomes true. If informal firms are relatively productive in the original problem, the threshold effect dominates (proposition 2) and firms have little to gain from becoming formal. This effect is reversed with the new specification. That is, growth implies higher penalties and that the incentive to become formal is strengthened. For large values of $A^i$ the firm is then able to escape such penalties by formalizing.
7 Conclusion

To the best of my knowledge, this is the first paper to use a dynamic model of profit maximizing firms to study investment incentives and formalization. In the model, there is a basic dynamic trade-off: On the one hand firms have an incentive to invest and grow in order to be able to reap the benefits from formalization. On the other hand, this may prove too costly and firms therefore remain informal. Formalization costs are taken at face value and the impact on investment paths and formalization of informal firms is studied.

The model used is consistent with recent evidence on marginal returns to capital at different firm sizes (McKenzie and Woodruff, 2006) Observing high returns for the smallest firm sizes but much lower returns for somewhat larger firms, the authors argue that entry non-convexities are unlikely to be a barrier to firm growth. Non-convexities such as formalization costs, however, relevant at larger firm sizes, may act as a barrier to firm growth.

The model generates a number of predictions. When aggregating over firms that differ along an ability dimension, the long-run firm size distribution exhibits a range of small firms and a range of larger firms but also a "missing middle", much in line with actual firm size distributions observed in developing countries (Bigsten et al. 2004, Tybout, 2000).

Another prediction, which in principle could be tested empirically, is that firms which are run by more able entrepreneurs formalize at smaller firm sizes.

With respect to policy, the paper offers two alternative interpretations on how policy vis-a-vis informal firms should be viewed. If the policy maker is a benevolent government, a policy designed to incentivize firms to become formal, should be conducted with a "carrot and stick" approach: neither too mild nor too tough. Such a policy will make an accumulation- and a threshold effect balance and maximize the amount of firms that become formal.

Finally, an interesting theoretical result is that an "informality trap" can result in a model with only a non-convexity but without a credit constraint. This contrasts with standard neoclassical theory, where non-convexities alone should not affect long run allocations. Economic agents could simply borrow to overcome such hurdles (Banerjee, 2001; McKenzie and Woodruff, 2006). In the present model it is the combination of adjustment costs of investment and the formalization cost that make initial capital matter for long-run capital distributions.
8 References


World Bank, 2009A. http://www.doingbusiness.org/

9 Appendix

9.1 Proof of proposition 1

Define the auxiliary capital level \( k_{\text{formalization}}^T \) as 

\[
A^i (\delta + \rho) + \delta A^f \overline{z} \delta (\delta + \rho)(2 \delta + \rho).
\]

This is the (hypothetical) level of capital that a firm would reach at \( T = \infty \) if it followed the formalization investment path forever. Using this expression, we can write the capital stock a firm reaches at the time of formalization, which is the LHS of expression (11), as

\[
k_{\text{formalization}}^T = \left( k_0 - k_{\text{informal}} \right) e^{-\delta T} + \left( k_{\text{informal}} - k_{\text{formalization}}^\infty \right) \delta e^{-(2 \delta + \rho) T}.
\]

As long as \( k_0 \leq k_{\text{informal}} \), \( k_{\text{formalization}}^T \) is increasing in \( T \) and converges to \( k_{\text{formalization}}^\infty \) as \( T \to \infty \). By equating \( k_{\text{formalization}}^\infty \) with \( k_F \) and solving for \( F \), the formalization criterion in proposition 1 is obtained.

For \( k_0 > k_{\text{informal}} \), the capital stock \( k_{\text{formalization}}^T \) reaches a maximum value of

\[
k_{\text{formalization}}^\infty + \left( k_0 - k_{\text{informal}} \right) e^{-\delta T} + \left( k_{\text{informal}} - k_{\text{formalization}}^\infty \right) \delta e^{-(2 \delta + \rho) T}.
\]

The obtainable capital stock is thus larger than \( k_{\text{formalization}}^\infty \) and depends on the initial stock of capital. As a result, the firm can face a higher formalization fee and still optimally choose to formalize. Comparing this capital stock with \( k_F \) and solving for \( F \), the second part of the proposition is obtained (with \( G(k_0) = \left( \frac{A^f - A^i}{\rho} \right) (\delta + \rho) \left( k_0 - k_{\text{informal}} \right) (2 \delta + \rho) \left( k_{\text{formalization}}^\infty - k_{\text{informal}} \right)^\delta / (\delta + \rho) \)).

9.2 Solving for the time allocated to hiding

The first order condition \(-\theta_j A^f - p(l) x \theta_j A^f (1 - l) + x \theta_j A^f (1 - p(l)) = 0\) gives the following solution:

\[
l^* = \frac{2 - 4x + 5x^2 - 2 \sqrt{(x - 1)^2 (1 - 2x + 4x^2)}}{9x^2}
\]

\[
p^* = \sqrt{T^*}
\]

\(A_j = \theta_j A^f h(x)\) where \( h(x) = 1 - l^* - x (1 - l^*) (1 - p^*)\)

Except for having \( h(0) = 1 \) and \( h(1) = \frac{2}{3\sqrt{3}} \) It can be shown that \( h'(x) < 0 \) and \( h''(x) > 0 \)
9.3 Penalties that increase in $k_t$

Choose $i_t$, $T$ to

$$\max \left[ T \left( A^T k_t \left( 1 - \frac{\beta}{2} k_t \right) - \frac{z}{2} i_t^2 \right) e^{-\rho t} dt + \int_T^\infty \left( A^T k_t - \frac{z}{2} i_t^2 \right) e^{-\rho t} dt - Fe^{-\rho T} \right]$$

s.t. $k_t = i_t - \delta k_t$ and $k(0) = k_0$

The solution to this problem is:

$$i_t = \delta \frac{C_2}{C_1} + C_4 (\delta + r_1) e^{r_1 t} + C_5 (\delta + r_2) e^{r_2 t}$$

$$k_t = \frac{C_2}{C_1} + C_4 e^{r_1 t} - C_5 e^{r_2 t}$$

where

$$C_1 = \frac{\beta}{z} + \sigma^2 + \delta \rho, \ C_2 = \frac{A_i}{z}$$

$$r_1 = \frac{\rho}{2} + \sqrt{\left( \frac{\rho}{2} \right)^2 + C_1}, \ r_2 = \frac{\rho}{2} - \sqrt{\left( \frac{\rho}{2} \right)^2 + C_1}$$

$$C_3 = \frac{1}{z} \left( \frac{1}{2} C_1 (\delta + r_1) e^{r_1 T} - (\delta + r_2) e^{r_2 T} \right)$$

$$C_4 = C_3 \left( C_1 F - \delta C_2 z + e^{r_1 T} (C_2 - C_1 k_0) (\delta + r_2) z \right)$$

$$C_5 = C_3 \left( -C_1 F + \delta C_2 z - e^{r_1 T} (C_2 - C_1 k_0) (\delta + r_1) z \right)$$