Optimal VAT Threshold: Official vs. Effective Enforcement

Jonathan Goyette*
Université de Sherbrooke

February 24, 2012

Abstract

Significant efficiency gains are available when there is a gap between official and effective enforcement of a value added tax (VAT). Using a unique dataset on Ugandan firms, I show that audits for taxes are effectively based on the number of employees rather than the official VAT threshold, which is in terms of sales. Based on the empirical evidence, I build a model of firms growth with entry and exit. Entrepreneurs evade part of their tax liabilities and, when audited, bargain with tax officials to keep some of the surplus from evasion in exchange of a bribe. The model is calibrated using the Ugandan data. The model does a reasonable job at replicating some features of the data that are not explicitly targeted and serves as a meaningful tool to conduct counterfactual experiments. The efficiency loss associated with evasion and corruption is of the order of 45% in Uganda. There is also a non-negligible gain in productivity per worker of 16% from enforcing the official VAT threshold based on the level of sales rather than the effective threshold based on the number of employees.

JEL codes: C15, E27, H26, L11, O16, O43
Keywords: Value Added Tax, Evasion, Corruption, Firm’s Growth, Size Distribution, Simulation

*Corresponding author: Jonathan Goyette; Address: 2500 Boulevard de l’Université, Sherbrooke (Quebec), J1K 2R1; tel: 1-819-821-8000 ext.62321; fax: 1-819-821-7364; email: jonathan.goyette@usherbrooke.ca
1 Introduction

Nowadays, more than 130 countries raise a good part of their tax revenues through a value added tax (VAT), whereas only a handful of countries were using a VAT fifty years ago (Keen and Lockwood, 2010). However, the degree of success of the VAT varies greatly. Least developed countries (LDCs) have particularly been prone to non-compliance. For example, the VAT almost collapsed in Uganda around 1995. Keen and Mintz (2004) suggest that this near-failure is due to the abrupt increase in the threshold. Using a unique dataset on Ugandan firms, I show in this paper that the implications are not just in terms of taxation revenues. I observe a mismatch between official and effective enforcement of the VAT threshold which generates non-negligible misallocation of resources and efficiency losses in Uganda. The sizeable effect of a switch from effective to official enforcement of the VAT threshold is large enough to account for some of the cross-country variation in productivity per worker (Hall and Jones, 1999).

It is a well-known fact that tax administrations in LDCs target larger firms more intensively because tax collection is expected to be greater than enforcement costs (Gauthier and Reinikka, 2006). Indeed, in the Article IV Consultation produced by the IMF for Uganda in 1998, it is noted that the Ordinary Audit Section has a "target... to carry out 75 audits a month, 50% of which would be on the largest taxpayers". However, a clear statement on how one evaluates the size of these largest taxpayers is nowhere to be found. One would expect that firms with a level of sales above the VAT threshold would systematically be audited for taxes. However, the data do not exhibit a clear monotone relationship between sales and audit intensity. Interestingly, I do observe such a relationship between the number of employees and audit intensity. It thus seems that, even if the VAT threshold is officially expressed in terms of sales, its effective enforcement is based on the number of employees.

This discrepancy between the official and effective procedures has three potential consequences. First, if assessment to audit firms for their taxes is based on a dimension that is easier to observe than sales, say the number of employees, resources could be misallocated. An extensive literature argues that some entrepreneurs stay small in order to pass under tax officials’ radar (de Soto, 1989; Gauthier and Gersovitz, 1997) or that entrepreneurs of larger firms segment their activities into many small firms to avoid restrictions based on firm size (Onji, 2009). The main question addressed in this paper is whether the discrepancy between what is officially said and what is effectively done generates such a misallocation of resources, translating furthermore into productivity losses.

Using data on 143 countries over 25 years, Keen and Lockwood (2010) show that most countries that adopted a VAT have developed a more effective tax instrument by testing whether an optimizing government increases its revenue ratio (the ratio of revenue from all sources to GDP). The results however do not hold for sub-Saharan Africa.
A second consequence of this discrepancy is that entrepreneurs expecting to be audited less intensively are more likely to evade their tax duties. Another objective of the paper is to quantify the productivity loss associated with evasion. Finally, entrepreneurs caught evading may offer bribes to keep part of the surplus from evasion. The last objective of this paper consists in examining whether entrepreneurs’ bargaining power over bribes helps in mitigating the tax distortion.

Based on firm-level data, I show that these three consequences are observed in Uganda. The data reveal that there is a statistically significant gap in the density of firm size where size is measured as the number of employees. This gap in the size distribution of firms is also known as the missing middle and is typical of many LDCs. As well, Ugandan firms evade about half of their tax obligations and pay 2.4% of their sales in bribes to keep part of the surplus from evasion.

Based on Hopenhayn (1992a), I suggest a model of firms’ growth with a taxation environment similar to the one observed in Uganda. In the model, firms produce a homogeneous good using capital and labor. Entrepreneurs decide whether to enter and exit the market and are heterogeneous along two dimensions: transient productivity shocks and current capital stock. Entrepreneurs make two choices in each period. First, they choose labor based on their current productivity shock and current capital stock. Second, they choose capital investment for the following period based on their expectations over future productivity. Entrepreneurs face an official VAT threshold based on the level of sales as well as an effective threshold based on the number of employees. Entrepreneurs evade part of their tax liabilities, and during an audit, they negotiate a bribe with their assigned tax official to keep some of the surplus from evasion.

Under some simplifying assumptions, I solve the model analytically and derive some comparative statics. I show that a change in firms’ bargaining power over the surplus from tax evasion has an ambiguous effect on the size of bribes and that this depends on the magnitude of the elasticities of the inputs with respect to the bargaining power. A similar pattern for the effect of the tax rate on bribes is observed. These results allow interpreting some of the findings in the numerical section.

The full version of the model is then calibrated. Parameters for the regulation environment are estimated using firm-level information from Uganda, while technology parameters are based on estimates from the existing literature, namely Hopenhayn and Rogerson (1993) and Restuccia and Rogerson (2008). The main objective of the estimation strategy for the remaining struc-

---

2 Many papers in the development literature argue that LDCs exhibit too few medium firms as compared to developed economies (Liedholm and Mead, 1987; Little et al., 1987; Sleuwaegen and Goedhuys, 2002; Steel, 1993; Steel and Webster, 1992; Tybout, 2000).

3 The model presented in this paper differs from Goyette and Gallipoli (unpublished results) in that I examine the negotiation process over bribes.
tural parameters requires some key statistics to arise endogenously, i.e., without explicitly being targeted. If these non-targets compare well with the data in dimensions relevant to the counterfactual experiments, this lends much credibility to the policy analysis. Targeted moments are related to production and corruption. I estimate parameters for initial and transitory productivity shocks processes and entrepreneurs’ bargaining power over bribes with a simulated method of moments. The simulated data also exhibit patterns in firm size, age and shares of output which compare well with the actual data without being explicitly targeted. The estimation strategy differs from the one in Goyette and Gallipoli (unpublished results) who target a set of moments related to key financial parameters, something I do not explore in this paper. Another important difference is that Goyette and Gallipoli (unpublished results) examine the effect of financial frictions on efficiency, whereas the main contribution of this paper does not depend on the modelization of financial frictions.

Having established the ability of the model to replicate some relevant features of the data, I examine four scenarios. First, I examine the effect of bargaining power over bribes on the allocation of resources and productivity. The model is constructed in such a way that bribing serves as a useful second-best mechanism in an environment already fraught with corruption. Indeed, bribes buy entrepreneurs a tax rebate and this, the larger their bargaining power. However, I find that the elasticities of capital and labor with respect to bargaining power ambiguously alter the effect of a change in bargaining power on output and bribes paid. Specifically, capital is more inelastic than labor when it is chosen intertemporally. This has important implications on productivity per worker since a decrease in entrepreneurs’ bargaining power may actually have a positive effect on efficiency, i.e., it could actually diminish aggregate labor so much that productivity per worker could increase while aggregate output is falling. These results should be contrasted with Svensson (2003) who argues that entrepreneurs with more bargaining power, as measured by sunk costs, pay smaller bribes.

Second, I examine the efficiency loss associated with evasion. In a revenue-neutral analysis, extra tax revenues recouped from evasion are used to subsidize consumption and generate a 45% increase in productivity per worker.4

In a third set of experiments, I derive the main result of this paper and quantify the efficiency gains associated with a switch from the effective VAT threshold based on the number of employees to the official VAT threshold based on a level of sales. Productivity per worker could increase as much as 16% in this scenario. It is important to stress that the main experiment of this paper is

---

4Keen and Smith (2007) have looked at the case of the EU where "carousel fraud" contributed to losses in the order of 10% of the net VAT receipts in some of the member states. By contrast, Uganda presents an environment typical of an LDC where the levels of informality and corruption are higher than in high-income countries. Although I do not examine carousel frauds specifically, it is of interest to note that evasion and corruption generate revenue losses in the order of 50% of the VAT receipts in Uganda.
not just a thought experiment like the two previous counterfactual scenarios. It is an assessment of the actual efficiency loss associated with the effective implementation of a policy.

In a final set of experiments, I examine whether the interaction of financial frictions and the level of the VAT threshold has an impact on efficiency. This is done for two reasons. First, I check whether the main result of this paper is independent of the assumption made in the theoretical section that entrepreneurs must finance capital internally. Introducing fully functioning capital markets has no impact on the main result of this paper. Second, many papers argue that a lower turnover threshold increases administrative costs through the number of firms to be handled by the VAT. However, I show that a decrease in the official VAT threshold to its pre-1995 level increases productivity per worker by 6%. This is due to two reasons. First, entrepreneurs must choose their level of productive capital before the realization of their productivity shock. Second, large firms generate the bulk of the output. Thus, lowering the VAT threshold alleviates its distortive interaction with financial frictions as smaller firms contribute less to aggregate efficiency. When there are no financial frictions, the effect of decreasing the VAT threshold is negligible on output per worker but there is a significant amount of reallocation of labor across firms. Hence, there is a tradeoff between administrative costs and efficiency in the presence of an interaction between financial frictions and the level of the VAT threshold.

The main contribution of the paper consists in showing that the efficiency consequences of a discrepancy between the official and effective enforcement of a VAT threshold are quantitatively large. To the best of my knowledge, this exercise has not been undertaken in the literature. My results shed light on potential efficiency losses in other situations where such a mismatch exists. Olken (2007) examines the discrepancy between official reported costs and independent engineers’ estimates of a road project in Indonesia and finds that increasing government audits reduces significantly misreporting of expenditures. Yang (2008) uses preshipment inspections of imports from a cross-section of countries to show that hiring private firms to conduct these inspections generates a decline in falsification of documentation and a significant increase in duty collections on imports.

This paper also hints at some potential causes explaining the collapse of the Greek tax administration, which adopted the EU system of Value Added Taxes (Kaplanoglou and Newbery, 2003). Spiro (1993) and Karoleff et al. (1994) show that underground activity has increased after the introduction of the goods and services tax (GST) in Canada in 1991. Using 1994 Canadian data on the GST, Piggott and Whalley (2001) show that this shift in production from formal to

---

6These results hold also using the effective VAT threshold based on the number of employees.
7The question whether the increase in administrative costs outweigh the increase in efficiency remains open since I do not have data on administrative costs in Uganda.
informal and relatively inefficient home producers decreased welfare.  

This paper also fits in a growing body of literature on misallocation of resources and cross-country differences in output per worker (see Buera et al., 2009; Goyette and Gallipoli, unpublished results; Hall and Jones, 1999; Hsieh and Klenow, 2007). Restuccia and Rogerson (2008) emphasize the necessity to obtain better measures of specific distortions and evaluate their aggregate consequences. This is exactly what I do in this paper: I identify a specific distortion due to tax audit and measure its consequences in terms of misallocation of resources and efficiency.  

Lastly there are a few papers that have examined the case of the VAT in Sub-Saharan Africa. Emini (2000) shows that an imperfect VAT has substantially increased fiscal revenues and investment to the expense of households’ welfare in Cameroon. Boccanfuso et al. (2011) argue that a tighter fiscal base might be more progressive than a multiple-rate VAT in the case of Niger. My paper departs from these in several dimensions: I account for the dynamics of investment, entrepreneurs use two inputs for production and bribe tax officials to keep the surplus from tax evasion.  

The rest of the paper is structured as follows. Section 2 presents the empirical evidence used to build the model of Sections 3 and 4. I describe a dynamic model with heterogeneous firms in Section 3. I also solve in closed-hand form a simpler version of the model and derive comparative statics for some parameters of interest. Section 4 presents the calibration strategy, the numerical implementation of the model and policy experiments. Section 5 concludes.

2 Empirical Foundation

The empirical section serves as a justification for the modeling choices of the next sections. In brief, the evidence is:

1. Firms below and above the VAT threshold pay the VAT.
2. Firms below and above the VAT threshold do not differ significantly in terms of their ratio of taxes paid to sales.
3. Audit intensity for taxes increases with the number of employees rather than the level of sales.
4. There is a stark increase in the audit intensity at 30 employees where I also observe a gap in the density of firm size.

According to Emran and Stiglitz (2005), substituting a VAT for tariffs increases fiscal distortions and diminishes welfare. The authors argue that these effects are due to the subset of formal commodities that can be taxed by a VAT.
5. Firms evade on average half of their tax duties.

6. Most firms who pay bribes did so to evade taxes.

7. Most firms do not fill out for compensation of duties paid on imports, even those registered for the duty drawback scheme.

2.1 The Data

The Ugandan Private Sector Foundation and the World Bank collected the data in 1998. The sample of 243 firms is random and representative, covers three years starting in 1995 and surveys businesses from five economic sectors: commercial agriculture, agro-processing, manufacturing, tourism and construction; as well as five geographical areas: Kampala, Jinja–Iganga, Mbale–Tororo, Mukono and Mbarara. The survey focuses on firms’ activities, including production, investment, finance, regulation, infrastructure, taxation, corruption and labor market.\(^9\)

In what follows, firms are classified as being small when hiring 30 employees or less. Medium firms hire between 31 and 75 employees. Large firms have 76 employees or more. The reason for this classification will become obvious in what follows.\(^10\)

2.2 Empirical Evidence

**Official VAT threshold** Figure 1 shows that VAT amounts increase with the level of sales. There is a jump in taxes paid at the threshold, i.e., $50,000. However, using a t-test to examine whether this difference in taxes paid is significant, I cannot reject the null that the ratio of taxes paid to sales is the same for firms below and above the VAT threshold. I examine the possibility of self-selection around the official VAT threshold, using McCrary (2008) local density estimator.\(^11\) Examining different windows in terms of bin size and bandwidth around the official VAT

---

\(^9\)For a detailed discussion about the data and summary statistics, see Svensson (2003).

\(^10\)Other papers in the development literature also use this classification. See, for example, Gauthier and Reinikka (2006) or the World Bank’s Enterprise Surveys.

\(^11\)The estimator is an extension of the local linear density estimator. In a first step, one obtains a finely gridded histogram of the size variable (number of employees). In the second step, this histogram is smoothed using a local linear regression separately on each side of the cutoff. The midpoints of the histogram are treated as regressors, and the normalized counts of the number of observations per size bins are treated as the outcome variable. One uses a weighted regression where more weight is given to the nearest bins, i.e., using a triangle kernel. The density function is then estimated by looping over evaluation points of the size variable. This algorithm provides a detailed distribution of the size variable which can then be plotted to give graphical evidence of the discontinuity around the threshold. The estimator suggested by McCrary (2008) has the advantage over traditional histogram techniques or kernel density estimates of allowing for point estimation or inference. Moreover, if one were to use a kernel density function separately for points to the left and right of the point of discontinuity, the density
threshold, I do not find convincing evidence of self-selection (results available upon request).\textsuperscript{12}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.jpg}
\caption{VAT amounts vs level of Sales in USD}
\end{figure}

**Effective VAT threshold**  In Figure 2, I examine the relationship between the audit intensity and firm size where size is measured either in terms of sales (left panel in Figure 2) or the number of employees (right panel in Figure 2). Comparing the two panels in Figure 2, the relationship between audit intensity and size is not monotonically increasing in the case of sales. However, the relationship between audit intensity and number of employees does increase smoothly.

In the model below, I introduce an audit probability which is a function of the number of employees using the simplest polynomial to approximate the pattern shown here: a step function. The choice of the step is based on an analysis of the density of firm size. The rationale for this approach is that if entrepreneurs are aware of the sharp increase in audit intensity beyond an approximate number of employees, they will self-select into a smaller size below this sharp increase. I use McCrary (2008)'s local density estimator to identify the effective VAT threshold in terms of number of employees. Table 1 reports a sensitivity analysis on the log difference in intercepts using different cutoff values between 10 and 35 employees. One clearly sees that the estimator suggested by McCrary (2008) is still better suited for this task since it corrects boundary bias at the right and left of the cutoff (Marron and Ruppert, 1994). Moreover, among non-parametric methods showing good performance at boundaries, the local linear density estimator is the simplest (McCrary, 2008).

\textsuperscript{12}However, even if there was self-selection around the official VAT threshold, this would not affect the main result of the paper. What I quantify in this paper is the effect of alleviating self-selection due to the effective VAT threshold, leaving the debate over the quantitative effect of self-selection around the official VAT threshold open.
largest and most significant log difference between intercepts occurs at 30 employees. The left panel of Figure 3 depicts the density estimate for firm size in the actual data. The vertical line on the graph represents the cutoff at 30 employees.\textsuperscript{13} I interpret this as evidence that entrepreneurs' labor choices are distorted because tax officials target firms above 30 employees for an audit.

![Figure 2: Audit intensity vs Firm Size](image)

Table 1: Comparison of Size Thresholds for the Density of Firm Size

<table>
<thead>
<tr>
<th>Size threshold (Nb. of Emp.)</th>
<th>Bandwidth (Nb. of Emp.)</th>
<th>Log Difference in Intercepts</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>29.0</td>
<td>-0.13</td>
<td>0.24</td>
</tr>
<tr>
<td>15</td>
<td>32.2</td>
<td>-0.62</td>
<td>0.25</td>
</tr>
<tr>
<td>20</td>
<td>35.7</td>
<td>-0.53</td>
<td>0.26</td>
</tr>
<tr>
<td>25</td>
<td>41.2</td>
<td>-0.73</td>
<td>0.28</td>
</tr>
<tr>
<td>30</td>
<td>47.6</td>
<td>-0.97</td>
<td>0.31</td>
</tr>
<tr>
<td>35</td>
<td>51.7</td>
<td>-0.51</td>
<td>0.31</td>
</tr>
</tbody>
</table>

The automatic selector from McCrary (2008) suggests the use of a bin size of 22 employees and a bandwidth of 48 employees. Given that a bin size of 22 employees does not convey a good graphical representation of the density around the threshold, I have use a bin size of 5 employees in Table 1 and Figure 3. These results are robust to choices of binsize between 5 and 50 employees. The choice of bandwidth however does affect the performance of the estimator. Table 2 reports the log difference in intercepts and the associated standard error for a cutoff of

\textsuperscript{13}The results presented here remain unchanged if I use an automatically selected bandwidth of 48 employees (more on this below).
30 employees and a bandwidth ranging between 25 to 50 employees. As can be inferred from Table 2, the results are robust for bandwidth choices of 35 employees and above.

<table>
<thead>
<tr>
<th>Bandwidth (Nb. of Emp.)</th>
<th>Log Difference in intercepts</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-.36</td>
<td>.48</td>
</tr>
<tr>
<td>30</td>
<td>-.43</td>
<td>.42</td>
</tr>
<tr>
<td>35</td>
<td>-.64</td>
<td>.38</td>
</tr>
<tr>
<td>40</td>
<td>-.82</td>
<td>.34</td>
</tr>
<tr>
<td>45</td>
<td>-.93</td>
<td>.32</td>
</tr>
<tr>
<td>50</td>
<td>-.99</td>
<td>.30</td>
</tr>
</tbody>
</table>

Finally, the difference in audit probability between small firms (30 employees or less) and larger firms (31 employees or more) is statistically significant at 30 employees using a t-test for groups with unequal variances. For obvious reasons, the significance of this result peters out as one increases the number of employees above 30 employees.\(^{14}\)

**Evasion and Corruption** The dataset contains information on the main taxes paid by Ugandan businesses: the value added tax (VAT), the corporate income tax (CIT), the National Social Security Fund (NSSF) levy, the excise tax, the withholding tax, import duties, the presumptive tax on small businesses and the local property tax. Gauthier and Reinikka (2006) report official rates and exemption conditions for each of these taxes. One of the main sources of tax exemptions is the 1991 Investment Code, which provides exemptions to large investors.

In a related paper (Azam et al., 2004), I construct a proxy for tax obligations using the Ugandan Tax Code and firms’ information on sales, size and exemptions. This constructed variable represents the amount of taxes officially owed by firms as opposed to what firms actually pay for the main taxes enumerated above. Subtracting taxes paid from tax obligations, we obtain an evasion surplus for each firm. Tax obligations represent 14% of sales in 1997 after exemptions. Taxes paid represent 7% of sales value. Overall, about half of the tax obligations disappear through evasion.

For the data on corruption, Reinikka and Svensson (2006) show that with appropriate survey methods and interview techniques, quantitative micro-level data on corruption can be obtained. As described in Svensson (2003), there are five key procedural components to collecting reliable

\(^{14}\)Noting that efficiency losses are smaller the lower the efficiency VAT threshold (as shown below in Table 7), one could interpret the results of the numerical section as an upper bound of the effect on efficiency of switching from the effective to the official VAT threshold.
information on bribes, which were used to gather the data. First, the main private sector organization (Uganda Manufacturers’ Association) carried out the survey. Relying on an organization that firms trust allowed circumventing deep-rooted distrust in the public sector. Second, information on bribe amounts was obtained indirectly. Respondents were asked to estimate the typical bribe payment a firm in their line of business would pay each year to deal with public officials, in customs, taxes, licences, regulation, etc.\textsuperscript{15} Third, questions related to corruption were asked at the end of the interview when trust and credibility were established. Fourth, there were multiple questions on corruption in different sections. This allows for numerous cross verifications. Lastly, there were at least two visits to each firm by one or two enumerators.

About 72\% of the firms in the sample report paying bribes. Out of the 67 that did not answer the bribe question, 27 did not answer other sensitive questions on costs, sales, etc. For the remaining 40 firms, Svensson (2003) shows that responders and non-responders do not differ significantly in terms of observables. Among the 176 firms that pay bribes, 48\% admit doing so to reduce tax obligations. The average bribe per employee is $71, and bribes represent 2.4\% of sales value.\textsuperscript{16}

\textbf{Compensation} The dataset provides information about firms compensation on duties paid on imports. On average, 28\% of firms’ inputs are imported. For firms below the official VAT threshold, this percentage decreases by half. However, when asked whether they have filled out for compensation from the duty drawback scheme 80\% of the sampled firms (and all firms below the VAT threshold) said they did not. A simple back-of-the-envelop calculation indicates why this is so. Suppose a firm has a level of sales just below the official VAT threshold. Input costs are $25,000 for this firm. The share of its imports in inputs is 14\% and the official tax rate for the VAT is 17\%. Compensation for this firm is thus approximately $595. In contrast, the annual workers wage in the dataset is $1,096. Since an accountant would command a higher salary, it is obvious that the costs from filing the compensation papers outweigh the gains. As a consequence, I do not include compensation on imports in the decision of entrepreneurs in the theoretical model although it can easily be amended to include this feature.

\textsuperscript{15}The question was: “Many business people have told us that firms are often required to make informal payments to public officials to deal with customs, taxes, licenses, regulations, services, etc. Can you estimate what a firm in your line of business and of similar size and characteristics typically pays each year?”

\textsuperscript{16}Tax obligations per employee are higher for medium firms which owe an average of $4,526 as opposed to $1,548 for small firms and $3,313 for larger firms (Goyette, Unpublished results). In terms of taxes paid per worker, small firms pay a third of what they owe, medium firms a bit less than half of their tax obligations, and large firms evade a fifth of their tax liabilities after exemptions. Medium firms are paying more bribes per employee ($127) than other firms ($47 for small firms and $73 for large firms). See Gauthier and Reinikka (2006) for more details on tax evasion and exemption in Uganda.
3 Theory

3.1 The Model

In this section I describe the basic elements of the model. The economy consists of entrepreneurs, workers and bureaucrats.

3.1.1 Entrepreneurs

Entrepreneurs are risk neutral and produce a homogeneous consumption good. In each period, entrepreneurs make hiring and production decisions based on their current productivity shock and current capital stock, as well as the probability of being audited by a tax official. Entrepreneurs who meet with a bureaucrat bargain over a bribe to keep part of the surplus from evasion.

All entrepreneurs on the market then decide their level of investment for the next period based on their expectations over productivity shocks. Entrepreneurs can only finance their investment through cash flow and invest in their own firm. Entrepreneurs who make no investment exit at the end of the period. Every period a fraction of firms exit and are replaced by new entrants at the beginning of the next period.

Entrepreneurs are heterogeneous across two dimensions. Upon entry in the market, each entrepreneur is endowed with a productivity parameter $z_t \in [z_{low}, z_{up}]$, drawn from an initial distribution function $\nu$. Productivity shocks then follow a transition function $g(z_{t+1}, z_t)$. Capital stocks $k_t \in (0, k_{up}]$ also differ across entrepreneurs because of previous investment history.

Since this paper is only concerned with stationary equilibria, I assume that entrepreneurs are rational and that the prices they anticipate are the prices that realize in equilibrium. The price of output is $P$ and the wage is $w$. In each time period $t$, an entrepreneur allocates his profits, $\pi_t$, between consumption, $c_t$, and capital investment, $k_{t+1}$. His budget constraint is:

$$c_t + k_{t+1} = \pi_t$$

where consumption and investment are assumed to be strictly greater than zero. Profits are

\footnote{This can be interpreted as a situation where the market for renting capital is not functioning because of conditions typical of LDCs, e.g., high default probabilities, usuary lending rates from private money lender, etc. (see Banerjee and Duflo, 2005). There are evidence in the data that small and medium firms do not have access to credit in Uganda and must rely solely on internal financing. Indeed, 53\% of the sample of firms did not request a loan during the period covered by the survey because either the collateral or the interest rate was too high. Goyette and Gallipoli (unpublished results) examine the effect of financial frictions on efficiency in Uganda. I check the implication of this assumption in the numerical section by examining the effect of introducing capital markets in the analysis. The model can easily be amended to have an integrated framework that allows for different levels of financial development.}
given by:

$$\pi_t = PF(z_t, k_t, l_t) - w_t l_t + (1 - \delta)k_t - \Omega$$  \hspace{1cm} (2)

where $l_t$ represents current labor and $\delta$ the depreciation rate of capital. The production function is $F(z, k, l) = zk^\gamma l^\eta$, which exhibits decreasing returns to scale, i.e., $\gamma + \eta < 1$. From a theoretical perspective, this allows pinning down the size of each specific firm. However, there is also an empirical justification for this choice. Basu and Fernald (1997) have estimated total returns to scale for a typical industry in the US, using data on 34 industries. They find that $\gamma + \eta$ lies between 0.8 and 0.9, a fact that I use in the numerical section. Finally, $\Omega$ represents the expected costs of taxation:

$$\Omega = \tau_p(z_o)PF(z, k, l) + p_a(z_o) \{ qB(z, k, l) + (1 - q) [(\tau_o(z_o) - \tau_p(z_o))PF(z, k, l) + E] \}$$  \hspace{1cm} (3)

where official and effective tax rates are $\tau_o(z_o)$ and $\tau_p(z_o)$, respectively. Both are a function of the productivity threshold $z_o$ associated with the VAT threshold (official or effective depending on the situation). The audit probability, $p_a$, is also dependent on $z_o$. The level of corruption in the tax administration is represented by $q$. This variable varies between 0 (no corruption) and 1 (full corruption, i.e., tax auditors always accept a bribe to let entrepreneurs evade). $B(z,k,l)$ is a bribe negotiated over the surplus from evasion between entrepreneurs and tax officials (more on this below). Finally, $E$ is a fine a firm must pay if caught evading by an honest bureaucrat.

Thus, a risk neutral entrepreneur maximize his lifetime utility:

$$E_0 \sum_{t=0}^{\infty} (\varphi \beta)^t c_t$$  \hspace{1cm} (4)

where $\varphi$ is an exogenous probability of survival and $\beta$ is the discount rate. Given his current shock $z_t$ and current level of capital stock $k_t$ an entrepreneur chooses to operate according to three different statuses $s = \{us, c, ul\}$. These statuses are defined by the VAT threshold (official or effective). There is thus a proportion $us$ of unconstrained-small entrepreneurs who face low productivity shocks and have a size (as measured by sales or the number of employees depending

---

18 A fixed cost is an important feature of this class of models à la Hopenhayn and has several implications. First, the fixed cost acts as a selection mechanism in that it affects the productivity-threshold for exit. However, the effect is ambiguous and depends on the relationship between the shock process and the elasticity of profits with respect to the fixed cost (Hopenhayn, 1992). Since I already have one selection mechanism (the VAT threshold), it seems reasonable to set the fixed cost to zero and concentrate on the effect of the mechanism of interest. Second, I cannot estimate this fixed cost with the data. Finally, estimating the fixed cost with the model would yield one more degree of freedom to fit the model to data.

19 As argued in the empirical section, entrepreneurs do not re-collect credits on inputs. The model can easily be amended to include this feature.

20 This exogenous death process is necessary to avoid a drift of the distribution of firms from period to period (Hopendayn, 1992a).
on the situation under examination) below the VAT threshold. These are unconstrained by the VAT threshold. Some entrepreneurs on the verge of crossing the VAT threshold prefer to constrain their inputs at $l^c(z_0)$ and $k^c(z_0)$ in order to keep size below its binding level. They constitute a proportion $c$ of constrained entrepreneurs. These entrepreneurs constrained their input choices so as to pass under the bureaucrats’ radar. There is a proportion $ul$ of unconstrained-large entrepreneurs such that, given $k_t$ and $z_t$, it is optimal to have a level of sales or number of employees above the VAT threshold. For a given $k_t$, let $z^ul_t$ be the endogenous shock-threshold beyond which an entrepreneur decides to produce above the VAT threshold.

3.1.2 Workers

Workers are risk-averse and face a hand-to-mouth budget constraint $c_t \leq w_t n_t$ where $c_t$ is workers consumption and $n_t$ is labor supply. As in Hopenhayn and Rogerson (1993), workers choose employment lotteries and diversify idiosyncratic risks. In the context of LDCs, the diversification of risks stems from extensive family ties or micro-credit. Workers are jointly represented through one representative agent with preferences given by:

$$\sum_{t=1}^{\infty} \beta^t [u(c_t) - DN_t]$$

where $N_t$ is the fraction of employed workers and $D$ is a disutility parameter.

3.1.3 Tax Officials

Based on the empirical evidence, each entrepreneur evade his tax liabilities. Upon being audited, an entrepreneur offers a bribe to the tax official inspecting his books. Entrepreneurs’ bargaining power is denoted by $\theta$ and stems from an imperfect technology to tell on corrupt tax officials and get them prosecuted.

The outside option for an entrepreneur being caught evading is to repay what he has evaded plus a fine $E$:

$$PF(z, k, l) - wl + (1 - \delta)k - \tau_o PF(z, k, l) - E$$

(6)

When a bribe is paid the profits of a firm are:

$$PF(z, k, l) - wl + (1 - \delta)k - B(z, k, l) - \tau_p PF(z, k, l)$$

(7)

The outside option for a tax official is to be honest and force the entrepreneur to pay his taxes in full, plus a fine. In this case, the tax official only receives his wage $w_b$. A corrupt tax official may accept a bribe from an entrepreneur. In this case, the bribe is added to his wage but he faces a probability $\psi$ of paying a fine $A$.
The Nash-bargaining problem is:

\[
B(z, k, l) = \arg \max \left[ (\tau_o - \tau_p)PF(z, k, l) + E - B(z, k, l) \right]^\theta \left[ B(z, k, l) - \psi A \right]^{(1-\theta)}
\]  

(8)

Solving for the equilibrium bribe yields:

\[
B(z, k, l) = (1 - \theta) \left[ (\tau_o - \tau_p)PF(z, k, l) + E \right] + \theta \psi A
\]

(9)

The equilibrium bribe shares the same properties as the production function and thus exhibits the following properties: \( B_l > 0 \), \( B_{ll} < 0 \), \( B_k > 0 \) and \( B_{kk} < 0 \).

### 3.2 Equilibrium

#### 3.2.1 Notation

First I look at the problem of an incumbent entrepreneur, taking entry as given. The labor decision is static, and employment is set so that the marginal product of labor equals the equilibrium wage. I can thus solve for labor in terms of capital. Bellman’s equation for an entrepreneur conditional on his status is:

\[
21V_s(k, z) = \max_k \left\{ c^s + \varphi \beta CV_s(k', z') \right\}
\]

subject to the budget constraint and where \( CV \) is the continuation value:

\[
CV_s(k', z') = \int_{z'} \max [V^{us}, V^c, V^{ul}] dg(z', z)
\]

(11)

At the end of each period before making their investment decision, entrepreneurs decide whether to stay in or exit the market. Exit occurs when the continuation value \( CV \) is less or equal to zero. The individual exit decision, \( c.e. \), is defined by a pair \( (k^{ex}, z^{ex}) \) such that:

\[
ex_s(k, z) = \begin{cases} 
0 & \text{if } CV_s(k', z') > 0 \\
1 & \text{if } \text{otherwise}
\end{cases}
\]

(12)

The unconditional value function is the upper envelope of the conditional value functions of an entrepreneur and it is given by:

\[
V(k, z) = \max_{s \in \{ex, us, c, ul\}} V_s(k, z)
\]

(13)

In order to obtain a stationary equilibrium, one requirement is to equalize the mass of entrants to the mass of entrepreneurs exiting the market (Hopenhayn, 1992a). The easiest way to fulfil

\[21\text{In the quantitative model, the algorithm iterates over the semi-conditionals until joint convergence.}\]
this condition is to assume that exiting entrepreneurs are automatically replaced by an entrant in the next period.

Taking entry as given, the entrepreneur’s decision problem produces three decision rules: one for the optimal choice of labor, one for the optimal choice of future capital and the other for the optimal exit decision: \( L^d, K \) and \( EX \).

The representative worker’s problem is a static optimization problem in a stationary equilibrium:

\[
\max u(c) - DN
\]  

subject to the workers’ hand-to-mouth budget constraint. The solution to this problem is denoted as \( L^o \).

A concise way to describe heterogeneity across individuals is to use a probability measure defined on subsets of the individual state space (see Lucas and Prescott, 1971; Hopenhayn, 1992). Let \( \mu(k, z) \) be such a probability measure, summarizing the state of the industry in a given period of time.\(^\ast\) Let \( G(k, z; k', z') \) be a transition function mapping current states into future states, which summarizes the optimal decisions of incumbents after exit decisions have been made. The evolution of the distribution of firms is then given by:

\[
\mu'(k', z') = \int \xi(k_0, z') \nu dz + \varphi \int (1 - ex(k, z))G(k', z'; k, z)\mu(k, z)d(k \times z) \tag{15}
\]

where \( \int \xi(k_0, z') \nu dz \) is the measure of entrants which equals the measure of exiting firms given by:

\[
\int [\varphi ex(k, z) + (1 - \varphi)]\mu(k, z)d(k \times z) \tag{16}
\]

Equation 15 states that next period’s measure of firms is given by the number of new entrants and the number of incumbents transiting from their current state \((k, z)\) to a future state \((k', z')\).

Having described the evolution of the industry, the government budget constraint can be defined. Government expenditures \( G \) are financed through taxes on the consumption good net of tax revenues lost through evasion and administrative costs:

\[
G = \max_{z_0} \int \{\tau_p(z_o)PF + p_a(z_o) [q \theta \psi A + (1 - q)((\tau_o - \tau_p)PF + E)]\} \mu(k, z)d(k \times z) - b(z_o, \psi) \tag{17}
\]

where the first term in the curly parentheses represents taxes effectively paid, the second term represents fines for corruption, taxes recovered by tax auditors and fines for evasion, and the term after the integral is a convex function for administrative costs to audit firms and monitor corrupt

\(^\ast\)As is standard in the literature on firms dynamics, an industry is a continuum of firms which produce a homogeneous good (Hopenhayn,1992a).
bureaucrats which has the following properties $b_p > 0$ and $b_z < 0$, i.e., administrative costs are increasing in the monitoring technology and decreasing in the productivity level associated with the VAT threshold.\footnote{The first assumption on the costs of monitoring technology is standard. The second assumption on costs associated with the VAT threshold is consistent with Keen and Mintz (2004) who find that a higher threshold reduces administrative costs as it diminishes the number of contributors to be handled.}

I now have all elements to define a stationary equilibrium.

**Definition 1:** A stationary equilibrium is a set of prices $\{w^*\}$, a set of decision rules, $\{z_o^*\}$, $\{z_{ul}^*\}$, $\{z_{ex}^*\}$, $\{l^d\}$, and $\{l^o\}$, and a distribution of firms $\{\mu^*\}$ such that:

1. Decision rules are optimal:
   - (a) given $k$, there exists a $z_o^*$ such that $V^{us} = V^c$ for each entrepreneur;
   - (b) given $k$, there exists a $z_{ul}^*$ such that $V^c = V^{ul}$ for each entrepreneur;
   - (c) given $k$, there exists a $z_{ex}^*$ such that the exit problem is satisfied for each entrepreneur;
   - (d) $l^d$ and $k^*$ satisfy each entrepreneur’s labor and capital problems;
   - (e) $L^o$ satisfy workers’ problem;

2. $\mu^*$ is defined recursively by 15 and $\mu'^* = \mu^*$;

3. $z_o^*$ is such that the government budget balances.

4. Prices are market clearing: $L^o = L^d$.

### 3.2.2 Optimal Choice for Labor

The first order conditions with respect to labor yields:

\[
l = \left[ \lambda(z_o) \left( \frac{\eta}{w} \right) z k^\gamma \right]^{1/\eta} \tag{18}\]

where $\lambda(l) = 1 + \tau_p - p_0(z_o)(1 - q^\theta)(\tau_o - \tau_p)$. The choice over labor depends on the size of a firm (where size is measured with sales or the number of employees). Larger firms face higher expected costs from taxation due to a higher probability of being audited. This has the obvious effect of distorting labor choices, especially for entrepreneurs just on the verge of crossing the effective VAT threshold.
3.3 Analytical Example

In this section I present some analytical results based on a solution in closed-hand form. However, some simplifying assumptions are needed to obtain such a solution. These restrictions are:

1. The production technology is $F(z, k, l) = z^{1 - \gamma - \eta} k^{\gamma} l^{\eta}$ with $\gamma + \eta < 1$.
2. Capital depreciates fully after one period and there is no investment.
3. Markets for labor and capital are complete and prices are defined by $P = 1$ for output, $r = r(K)$ for capital and $w = w(L)$ for labor where $r$ and $w$ are strictly decreasing in $K$ and $L$ respectively.\(^{24}\)
4. Productivity shocks are independent and uniformly distributed between 0 and 1.
5. Enforcement is weak, i.e. $E = 0$ and $\psi A$ is such that all tax officials always accept a bribe, i.e. $q = 1$.
6. There is no entry and no exit.
7. The audit probability is zero for unconstrained-small firms and $b$ for unconstrained-large firms.

These assumptions render expressions for capital and the productivity threshold $z_{ul}$ analytically tractable.\(^{25}\) Moreover, note that another implication of the assumption of perfect capital market is a one-to-one relationship between the level of productivity and capital as well as labor. Therefore, the VAT threshold in terms of sales or the number of employees can be associated with a single productivity class $z^o$, which in turn can be associated with a unique level of capital and labor, i.e., $l^c$ and $k^c$. The analytical expressions for the variables of interest, namely decision rules over capital and labor for unconstrained and constrained firms as well as the shock-threshold $z^u$ are presented in the appendix.

3.3.1 Effect of a Change in the Bargaining Power of Entrepreneurs

I examine the effect of a change in the bargaining power of entrepreneurs over the surplus from evasion. Given the assumptions in this example, I am only interested in the effect of bargaining power on capital and labor for unconstrained-large firms:

\(^{24}\)This is as in Hopenhayn (1992a).
\(^{25}\)Relaxing assumptions 2, 3 or 4 requires computational methods to solve the model. This is done in the next section.
The equilibrium bribe is a function of both inputs $k^{ul}$ and $l^{ul}$. In turn, these are both functions of $\theta$, and thus:

\[
\frac{dk^{ul}}{d\theta} = \frac{\tau_o}{(1 - \gamma - \eta)} \lambda^{1 - \gamma - \eta} z \left[ \frac{(\gamma)^{1-\eta} \left( \frac{\eta}{w} \right)^\eta}{\gamma} \right]^{1/(1-\gamma-\eta)} > 0 \quad (19)
\]

\[
\frac{dl^{ul}}{d\theta} = \frac{\tau_o}{(1 - \gamma - \eta)} \lambda^{1 - \gamma - \eta} z \left[ \frac{(\gamma)^{\gamma} \left( \frac{\eta}{w} \right)^{1-\gamma}}{\gamma} \right]^{1/(1-\gamma-\eta)} > 0 \quad (20)
\]

The equilibrium bribe is a function of both inputs $k^{ul}$ and $l^{ul}$. In turn, these are both functions of $\theta$, and thus:

\[
\frac{dB}{d\theta} = \tau_o z^{1-\gamma-\eta} k^{\gamma} l^{\eta} \left\{ -1 + \frac{(1 - \theta)}{\theta} [\gamma \varepsilon_{k^{ul}\theta} + \eta \varepsilon_{l^{ul}\theta}] \right\} + \psi A \geq 0 \quad (21)
\]

where

\[
\varepsilon_{k^{ul}\theta} = \frac{\theta}{k^{ul}} \frac{dk^{ul}}{d\theta} > 0 \quad (22)
\]

and

\[
\varepsilon_{l^{ul}\theta} = \frac{\theta}{l^{ul}} \frac{dl^{ul}}{d\theta} > 0 \quad (23)
\]

A change in bargaining power has ambiguous effects on the size of the bribe. Indeed, as can be noted from Equation 21, the first term on the RHS (call it RHS1) is negative, while the second and third terms (call them RHS2 and RHS3) are positive. When enforcement is greater than zero bureaucrats pass on the expected fine, $\psi A$, to entrepreneurs and this, the higher the bargaining power of an entrepreneur. In the limit, when an entrepreneur has all bargaining power, it must assume the total expected fine faced by the tax official who gains nothing from the negotiation. The proof of the following proposition is in the appendix:

**Proposition 1** If the elasticities of the inputs with respect to the bargaining power are large, the bargaining power of a firm has to be high to reduce the bribe. Conversely, when the elasticities are small, entrepreneurs do not change their allocations by much and have more leverage to negotiate a lower bribe.

**Proof:** Appendix.

The effect of a change in bargaining power on the shock-threshold $z^{ul}$ is:

\[
\frac{dB}{d\theta} = \tau_o z^{1-\gamma-\eta} k^{\gamma} l^{\eta} \left\{ -1 + \frac{(1 - \theta)}{\theta} [\gamma \varepsilon_{k^{ul}\theta} + \eta \varepsilon_{l^{ul}\theta}] \right\} + \psi A
\]

and then multiply the second term on the right hand side (RHS) by $\frac{d}{d\theta}$ to get Equation 21.

\[
(1 - \gamma - \eta) \left( z^{ul} \right)^{1-\gamma-\eta} \frac{dz^{ul}}{d\theta} A
\]

\[
26\text{To obtain this equation, first derive B with respect to } \theta:\n\frac{dB}{d\theta} = \tau_o z^{1-\gamma-\eta} k^{\gamma} l^{\eta} \left\{ -1 + \frac{(1 - \theta)}{\theta} [\gamma \varepsilon_{k^{ul}\theta} + \eta \varepsilon_{l^{ul}\theta}] \right\} + \psi A
\]

\[
27\text{To see this, take the derivative of Equation 38 with respect to } \theta:\n(1 - \gamma - \eta) \left( z^{ul} \right)^{1-\gamma-\eta} \frac{dz^{ul}}{d\theta} A
\]
\[
d\frac{dz^{ul}}{d\theta} = \left[\psi A - \tau_o (z^{ul})^{1-\gamma-\eta} (k^{ul})^\gamma (l^{ul})^\eta \right] \left(1 - \gamma - \eta \right) (z^{ul})^{-\gamma-\eta} \Lambda \geq 0
\]

where \( \Lambda = \{\lambda (k^{ul})^\gamma (l^{ul})^\eta - (k^c)^\gamma (l^c)^\eta\} \)

An increase in bargaining power has ambiguous effects on the shock threshold. Note however that no bribes are exchanged in equilibrium if \( \psi A \geq \tau_o (z^{ul})^{\lambda} (k^{ul})^\gamma (l^{ul})^\eta \). This implies that, in a situation where bribes are exchanged, the share of unconstrained-large firms increases with an increase in bargaining power which is as expected.

I find a similar set of results for the effect of a change in the official tax rate, \( \tau_o \). The results are in the appendix.

4 Quantitative Model

In this section I revert to the original, dynamic model with capital investment.

4.1 Calibration Strategy

The parameters for the fiscal environment are estimated using the Ugandan data while technology parameters are based on the existing literature. I use the model to estimate three structural parameters based on a minimum distance criterion:

\[
\hat{\phi} = \arg \min_{\phi} (m_s(\phi) - m_d)'^TW(m_s(\phi) - m_d)
\]

where \( W \) is the variance-covariance matrix from the data, \( m_d \) are moments from the data and \( m_s \) are moments from the simulated data obtained in the exact same way as those from the actual data.

Equation 25 is analytically intractable and I cannot rely on standard optimization tools based on gradients to search for a minimum. This is due to the non-convexity introduced by modelling the VAT threshold. I use Simulated Annealing, a minimization algorithm that proceeds by random search.

\[
(z^{ul})^{1-\gamma-\eta} \left\{ \tau_o (k^{ul})^\gamma (l^{ul})^\eta + \lambda \left[ \gamma (k^{ul})^\gamma-1 (l^{ul})^\eta \frac{dk^{ul}}{dr} + \eta (k^{ul})^\gamma (l^{ul})^\eta-1 \frac{dl^{ul}}{dr} \right] \right\}
\]

\[ -w \frac{dl^{ul}}{dr} - r \frac{dk^{ul}}{dr} = \psi A \]

Using the equilibrium conditions for wages and the rental rate to eliminate the relevant terms in the above expression I get the equation in the text.
Table 3 presents the value and origin of all parameters used in the benchmark calibration. Note that a time period in the model is one year in the data.

4.1.1 Regulation Environment

For the sake of simplicity, I assume that all entrepreneurs face the same official and effective tax rates. The official tax rate for the VAT is 17%. However, based on information on exemptions, I find a ratio of tax obligations per sale in the data such that $\tau_0 = 0.14$. Furthermore, empirical evidence are such that the effective rate of taxes paid, $\tau_p$, is half the official rate. The official VAT threshold is $50,000$ based on the Ugandan tax code. Estimates of audit intensity are 53% for small firms and 71% for larger firms.\(^\text{28}\)

The bargaining power of entrepreneurs is estimated using the model. The targeted moment in Equation 25 is the average of the ratio of bribes per sales in data. I use the average bribe per employee observed in the data as an estimate for $\psi A$. Note that the average bribe per employee is approximately half a month of a tax official’s wage. The Article IV Consultation produced by the IMF for Uganda in 1998 estimates the average public wage at approximately $w_b = $1,800 in 1997.\(^\text{29}\) This entails a bribe per firm (for the average firm with 124 employees), which is about five times the yearly wage for tax officials. Hence, bribes need not be large from the vantage point of entrepreneurs to convince tax officials to let them go with the surplus from evasion.\(^\text{30}\)

Based on the Article IV Consultation, only 16% of the assessments were refunded through ordinary tax audits for the period from July to December 1997. I assume that the poor performance of the audits is due to the fact that all tax officials are potentially corruptible and I set in the benchmark $q = 1$. The report also states that compliance to fill out tax reports was low, implying incentives to do so where weak. I set the fine paid by entrepreneurs, $E$, to zero.

4.1.2 Technology

Following Restuccia and Rogerson (2008), I use $\gamma + \eta = .85$ and attribute parameter values according to capital and labor shares of income, i.e., 1/3 and 2/3, respectively.\(^\text{31}\) I assume a rate

\(^{28}\)Evidence show that the official VAT threshold and the audit intensity affect all taxes. This is why I use the ratio of taxes paid to sales and not just the ratio of the VAT to sales in the calibration.

\(^{29}\)The total public wage bill for 1997 in Uganda was 227 billions of Ugandan Shillings. The public work force was 124,664 employees.

\(^{30}\)This partly answers the questions of Gang et al. (1988) who ask: ”If public workers [in LDCs] suffer discrimination by wage, why is it that demand for such jobs is high?” My results also corroborates Lindauer et al. (1988) who describe the case of the Ugandan administration in the eighties where ”civil servants either had to diminish their ethic standards or perish in uprightness”.

\(^{31}\)Note that Soderbom and Teal (2004) find such an assignment of income shares for Ghana, a country similar to Uganda.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.97</td>
<td>Intertemporal discount factor</td>
<td>literature</td>
</tr>
<tr>
<td>D</td>
<td>1.22</td>
<td>Disutility of labor</td>
<td>assumption</td>
</tr>
<tr>
<td>$F(z_o)$</td>
<td>US $50,000</td>
<td>Official VAT threshold</td>
<td>Ugandan Tax Code</td>
</tr>
<tr>
<td>$l^c$</td>
<td>30</td>
<td>Effective VAT threshold</td>
<td>data</td>
</tr>
<tr>
<td>$\tau_0$</td>
<td>0.14</td>
<td>Official tax rate</td>
<td>data</td>
</tr>
<tr>
<td>$\tau_p$</td>
<td>$\tau_0/2$</td>
<td>Effective tax rate</td>
<td>data</td>
</tr>
<tr>
<td>$p^s_a$</td>
<td>0.53</td>
<td>Probability of audit small firms</td>
<td>data</td>
</tr>
<tr>
<td>$p^l_a$</td>
<td>0.71</td>
<td>Probability of audit large firms</td>
<td>data</td>
</tr>
<tr>
<td>$q$</td>
<td>1</td>
<td>Probability to accept a bribe</td>
<td>assumption</td>
</tr>
<tr>
<td>$\psi A$</td>
<td>71USD/yr-emp</td>
<td>Expected fine for tax officials</td>
<td>data</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.88</td>
<td>Firm’s bargaining power</td>
<td>estimated using model</td>
</tr>
<tr>
<td>$E$</td>
<td>0</td>
<td>Fine for entrepreneurs</td>
<td>assumption</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.07</td>
<td>Depreciation Rate</td>
<td>literature</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.28</td>
<td>Share of capital in production</td>
<td>literature</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.57</td>
<td>Share of labor in production</td>
<td>literature</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.98</td>
<td>Persistence</td>
<td>estimated using model</td>
</tr>
<tr>
<td>$\sigma^2_e$</td>
<td>0.03</td>
<td>White noise’s variance</td>
<td>assumption</td>
</tr>
<tr>
<td>$w$</td>
<td>$1,096/yr$</td>
<td>Wage</td>
<td>measured from data</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.9</td>
<td>Firms’ survival rate</td>
<td>literature</td>
</tr>
<tr>
<td>$n_z$</td>
<td>20</td>
<td>Number of productivity nodes</td>
<td>literature</td>
</tr>
<tr>
<td>$n_i$</td>
<td>16</td>
<td>Maximum initial productivity</td>
<td>estimated using model</td>
</tr>
</tbody>
</table>
of depreciation of $\delta = 0.07$.

As in Hopenhayn and Rogerson (1993), the price of output, productivity shock and capital stock enter multiplicatively in the production function. A high price is hard to distinguish from a high value of the current shock or capital stock. To deal with this issue, I make a few assumptions. First, the price of output is normalized to one. Second, the equation for the relative demand for labor is used to pin down the range of productivity shocks:

$$\frac{l_i}{l_j} = \left[ \frac{\lambda_i z_i}{\lambda_j z_j} \left( \frac{k_i}{k_j} \right)^{1-\eta} \right]^{1/\theta} \quad (26)$$

I use a capital grid with 250 log-spaced points where the lower bound is normalized to one. A sensitivity analysis is used to locate the upper bound: no firms should cluster at the upper bound on the capital grid. This provides the smallest possible capital grid which economizes on computational time. Note also that the grid is as in Hopenhayn and Rogerson (1993). Firm size ranges between 1 and 2000 employees in the Ugandan data. This generates a support for shocks which ranges from $z_{\text{low}} = 1$ to $z_{\text{up}} = 3.8$. Also following Hopenhayn and Rogerson (1993), I impose a grid for productivity shocks with a number of evenly spaced nodes, $n_z$, equal to 20.

Productivity shocks for new entrants are initially drawn from a uniform distribution, defined on a subset of the productivity domain of incumbent firms. The range of the initial distribution of productivity shocks is estimated using the model and denoted as $n_i$. Initial productivities thus range over the $n_i$ lowest points of the productivity grid. In order to identify this parameter I use the average number of employees as a target in equation (25). For firms that have been in the market for at least one period, productivity shocks follow an AR(1) process:

$$z_t = \rho z_{t-1} + \epsilon_t \quad (27)$$

where $0 < \rho < 1$ and $\epsilon_t$ is a white noise perturbation. I assume that the distribution of the white noise is a Normal with mean zero and variance $\sigma^2$. The AR(1) process is covariance-stationary with mean zero and variance $\sigma^2_z = \sigma^2_{\epsilon_t}/(1 - \rho^2)$.32 I use a method suggested by Kopecky and Suen (2010) to approximate the AR(1) process with a discrete grid. This involves constructing a grid for productivity shocks with $n_z$ points and an upper bound given by:

$$z_{\text{max}} = \sqrt{(n_z - 1)\sigma_z} \quad (28)$$

A transition matrix $g(z_{t+1}, z_t)$ is then computed using a parameter $p = 1 + \rho$. I estimate $\rho$ using the simulation-based method described above. I use the capital-output ratio as a target in equation 25 to identify this parameter.33 Finally, I pin down the value of $\sigma^2_{\epsilon_t}$ using equation 28. Given that $z_{\text{max}}$ is equal to 3.8, I use the estimated value for $\rho$ to back out $\sigma^2_{\epsilon_t}$ in equation 28.

---

32 It can be shown that if $\epsilon_t$ is normally distributed then $z_t$ is also normally distributed.

33 I present in the appendix the approach to estimate capital with the Ugandan data.
I make middle of the range choices for the survival rate of firms and the discount factor. For the survival rate of firms, $\varphi$, Roberts and Tybout (1997) estimate that the one-year exit rate for firms in Chile, Colombia and Morocco are 8.5%, 11.9% and 9.5%, respectively. I use a value of 10% for Uganda and set $\varphi = 0.9$. For the discount factor of entrepreneurs, Arellano, Bai and Zhang (2007 and 2008) use $\beta = 0.96$ for Ecuador and $\beta = 0.94$ for Bulgaria. Many papers based on US data use a discount rate around 0.98 (e.g. Atkeson and Kehoe, 2005). I use $\beta = 0.97$.

The wage is calculated based on the optimal labor equation 18 for the smallest firm in the model, i.e., with one employee, the entry level of capital and the lowest shock. The corresponding wage in the data is $1,096. I assume that workers are risk averse and that $u(c_t) = \log(c_t)$. The combination of the assumption of log-utility and a hand-to-mouth budget constraint implies that aggregate labor is inelastic in this model. To see this, take the first order condition of the workers’ problem. This yields $N = 1/D$. Using UN data for the employment-to-population for Uganda in 1997, I get $D = 1.22$. In the experiments below, I adjust the wage so that aggregate labor is kept as in the benchmark. I thus examine the effect on efficiency of reshuffling labor across productive units when there is a change in the structural environment.

In summary, I estimate three parameters using a simulated method of moments. These parameters are: $\rho$, $\theta$ and $n_i$. There are four targeted moments: (i) the average number of employees; (ii) the average bribe-output ratio; (iii) the average capital-output ratio; (iv) the share of medium firms. I thus have a vector of three parameters for $\zeta$ and a vector of four targeted moments. The reason for this overidentification is that a just-identified model underestimates the share of medium firms. This is mainly due to the stark depiction of the effective VAT threshold at 30 employees in the model. This is a disadvantage of using a polynomial with only one step. Adding the share of medium firms as a target improves the fit of the model. However, there is still one degree of freedom associated either with the share of small or large firms, leaving room for endogenous self-selection by the entrepreneurs of the model.

### 4.2 Model Fit

#### 4.2.1 Targets

Table 4 contrasts the simulated moments obtained from the calibration with the targeted moments in the data. The model does a reasonable job in fitting the actual data. All simulated moments lie well within a confidence interval of one standard deviation of the actual moments. The share of medium firms is slightly overestimated while the average capital to output ratio and the average number of employees are moderately underestimated. However, the average bribe to output ratio is right on target.

---

34Employment-to-population ratio for Uganda in 1997 for both sexes is 82%, 88% for men and 76% for women.
Table 4: Targets

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benchmark</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of employees</td>
<td>122</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>(259)</td>
<td></td>
</tr>
<tr>
<td>Average bribe/output ratio</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td></td>
</tr>
<tr>
<td>Average capital/output ratio</td>
<td>1.01</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>(3.1)</td>
<td></td>
</tr>
<tr>
<td>Share of medium firms</td>
<td>0.23</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Note: Std. dev. in parenthesis

4.2.2 Non-Targeted Statistics

This section establishes whether the model does a reasonable job at replicating some features of the data that were not explicitly targeted. Table 5 summarizes these results.

Table 5: Non-Targeted Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Size bins</th>
<th>Benchmark</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of firms</td>
<td>Small</td>
<td>.46</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>.23</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>.31</td>
<td>.32</td>
</tr>
<tr>
<td>Average Size</td>
<td>Small</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>57</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>334</td>
<td>334</td>
</tr>
<tr>
<td>Average Age</td>
<td>Small</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Maximum Age</td>
<td>All</td>
<td>84</td>
<td>74</td>
</tr>
<tr>
<td>Share of Output</td>
<td>Small</td>
<td>.06</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>.11</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>.83</td>
<td>.83</td>
</tr>
</tbody>
</table>

Firm Size Distribution The model generates an equilibrium size distribution of firms that compares well with the data as observed from Table 5. In terms of shares, small and large firms are underestimated, while medium firms are overestimated. Figures for the average number of
employees per firm generated by the model match the data pretty well for small and large firms. Medium firms in the model tend to be bigger than what is actually observed. This compensates for the fact that the share of large firms is underestimated while targeting the average number of employees.

Suggestive graphical evidence of the fit of the model is presented in the panels of Figure 3. The panels allow comparing the density of firm size from the actual and simulated data. The density is estimated using McCrary (2008)’s local density estimator.

The estimator allows testing for a discrete change in the density of firms’ sizes around the effective VAT threshold of 30 employees. The discontinuity in the density at 30 employees is highlighted by a vertical line in both panels of Figure 3. The difference in the intercepts in the data [benchmark] at that point is equal to -.99 [-1.0] and its standard deviation is 0.3 [0.32] with a p-value equal to zero. The model reproduces a clustering of firms below the effective VAT threshold. This evidence is consistent with the hypothesis that entrepreneurs refrain from growing beyond 30 employees knowing they will incur a rise in expected costs from taxation. This corroborates a similar finding in Onji (2009) who observes a significant rise in registration at the VAT treshold, indicating that Japanese firms are splitting their operations to avoid similar concerns as their Ugandan counterparts. As the theoretical model shows, sub-optimal input choices create this gap in the size distribution of firms labelled the ”missing middle” in the development literature (see Tybout, 2000; Liedholm and Mead, 1987; Little et al., 1987; Steel, 1993; Steel and Webster, 1992; Sleuwaegen and Goedhuys, 2002).35

**Firms age and share of output** Small firms in the model are younger than their counterparts in the data. However, the model does a good job at estimating the average age of medium and large firms with the specific feature that medium and large firms have approximately the same age. The age of the oldest firm is over-estimated by the model. The model overestimates the share in aggregate output of small firms to the expense of medium firms. However, it provides a good estimate of the share of aggregate output for large firms.

4.3 Counterfactual Experiments

The fit of the model seems reasonable in dimensions not targeted explicitly and relevant policy-wise. I thus conduct in this section a set of counterfactual experiments to examine the impact on efficiency and the size distribution of firms of 1) bribes, 2) evasion, 3) switching the VAT

---

35Fortin et al. (1997) offer an explanation for the missing middle in Cameroon based on the fact that the marginal cost of tax and regulation evasion increases with size. However, capital is absent from their model and this does not allow entrepreneurs to substitute capital for labor when facing a distortion on labor in order to minimize efficiency losses as presented here.
threshold from its effective to its official level, and 4) the interaction of financial frictions with the level of the official VAT threshold. Table 6 exhibits the results of experiments involving the former three issues. Table 7 presents the results related to the interaction between financial frictions and the official VAT threshold. All percentage changes presented in these tables are given relative to the numerical benchmark.

4.3.1 Bargaining Power

What is the effect of bribes in this model? I answer this question by setting the bargaining power of entrepreneurs ($\theta$) to zero. In Column 1 of Table 6, I examine the effect of reducing bargaining power without any other adjustments in the model. This can be considered as an examination of short-run effects of a change in bargaining power. As noted in Column 1 of Table 6, aggregate output falls by 8%. This is as expected: bargaining power buys a tax rebate to entrepreneurs in the model. The lower the bargaining power, the smaller the share of the surplus from evasion an entrepreneur keeps for himself. However, there is also a decrease in aggregate labor, which is larger in magnitude (12%) than the fall in aggregate output. This is mainly due to a decrease of 7% in the share of large firms. Overall, it implies that output per worker increases when bargaining power is set to zero and the wage is not adjusted.

Let’s use some analytical results from the theoretical section to understand what is going on. When capital markets are fully functioning as in the analytical example, one can easily show that the elasticity of capital with respect to the bargaining power is equal to the elasticity of labor, i.e., $\epsilon_{k\theta} = \epsilon_{l\theta}$\textsuperscript{36} However, in a situation where capital is chosen before the realization of the shocks, like in the numerical benchmark, capital cannot be adjusted optimally. This implies that

\textsuperscript{36}Proof in appendix.
the response of capital to a change in bargaining power will be less elastic than the response of labor under incomplete capital markets. This can be noted from Column 1 of Table 6: aggregate capital changes by 7% while the change in aggregate labor is 12%. Hence, a larger share of the adjustment in output due to the change in bargaining power takes place through labor than capital.

In Experiment 2, I adjust the wage so that aggregate labor remains as in the benchmark. Recall: log-utility for workers implied an inelastic supply of labor. This second experiment considers the long-term effect of a change in bargaining power. As observed in Column 2, the wage must be reduced by 3% to keep aggregate labor as in the benchmark. Now the reallocation of labor goes in the opposite direction: there is a gain of 5% in the share of large firms at the expense of small and medium firms. This generates an increase in output per worker of 2%. Here, the gain in output per worker is due to the offsetting effect on labor of decreasing the wage. Given that aggregate labor must be kept as in the benchmark, entrepreneurs take advantage of the lower wage by acquiring more capital. Although gains in aggregate output appear to be one for one with gains in aggregate capital, the reshuffling of labor across firms contributes also to this gain in productivity.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corruption</td>
<td>θ = 0</td>
<td>I = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>τc</td>
<td>—</td>
<td>—</td>
<td>-14</td>
<td>—</td>
<td>—</td>
<td>-3</td>
</tr>
<tr>
<td>Δwage</td>
<td>—</td>
<td>-3</td>
<td>18</td>
<td>—</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>ΔShareSmall</td>
<td>3</td>
<td>-2</td>
<td>6</td>
<td>-23</td>
<td>-12</td>
<td>-11</td>
</tr>
<tr>
<td>ΔShareMedium</td>
<td>4</td>
<td>-1</td>
<td>-53</td>
<td>33</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>ΔShareLarge</td>
<td>-7</td>
<td>5</td>
<td>30</td>
<td>10</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>ΔAgg.Capital</td>
<td>-7</td>
<td>2</td>
<td>137</td>
<td>26</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>ΔAgg.Output</td>
<td>-8</td>
<td>2</td>
<td>45</td>
<td>32</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>ΔAgg.Taxes</td>
<td>-8 [57]</td>
<td>2 [74]</td>
<td>0</td>
<td>32</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>ΔAgg.Bribe</td>
<td>73 [-100]</td>
<td>92 [-100]</td>
<td>-100</td>
<td>34</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>ΔAgg.Labor</td>
<td>-12</td>
<td>—</td>
<td>—</td>
<td>32</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: Change in aggregate taxes and bribes for q=0 in brackets.
4.3.2 Evasion

In this section, I examine the impact of evasion.\textsuperscript{37} The analysis is very similar to the previous experiments, except that now \( q = 0 \) instead of \( \theta = 0 \). Note that these two variables have the same impact on expected costs from taxation in equation 18: 
\[
\lambda(l) = 1 + \tau_p - p_a(z_o)(1 - q\theta)(\tau_o - \tau_p).
\]
For this reason, I do not repeat the exercise with and without wage adjustment: the results are the same as in columns 1 and 2, except for changes in taxes and bribes. When \( \theta = 0 \), entrepreneurs transfer all surplus from evasion to corrupt bureaucrats, while the surplus from evasion is recovered by the government when \( q = 0 \). Hence, the values in brackets in columns 1 and 2 are changes in taxes and bribes when \( q \) is set to zero. Then, the results of a revenue-neutral analysis are presented in Column 3 in Table 6. I introduce a subsidy on consumption, \( \tau_c \), to keep the level of taxes collected as in the benchmark.\textsuperscript{38} The wage is also adjusted to keep aggregate labor as in the benchmark.

In the experiment behind the results of Column 3 of Table 6, there is a doubling in expected costs from taxation from the vantage point of entrepreneurs. This has two effects: 1) entrepreneurs scale down their operations and 2) the aggregate amount of taxes collected by the government increases. The extra revenues from taxation can be used to subsidize consumption (\( \tau_c = -.14 \)). This stimulates aggregate demand and entices entrepreneurs to scale up their operations. As a consequence, the share of large firms increases by 30% relative to the benchmark. The wage also increases to keep aggregate labor constant. This probably explains why the share of small firms increases by 6%. Clearly, the missing middle is exacerbated in this experiment. The jump in expected taxes at the effective VAT threshold is more important than in the benchmark. Nonetheless, evasion generates a loss in output per worker that could be as large as 45% as can be noted from Column 3 in Table 6.

4.3.3 Official vs. Effective Threshold

This section answers the main question of the paper: what is the gap in efficiency between the effective VAT threshold based on the number of employees and the official VAT threshold based on sales? In Experiment 4, I examine the switch from effective to official VAT threshold without any other adjustments in the model. Firms below the official VAT threshold face a probability of being audited equal to zero while the probability of being audited above the official VAT threshold is kept at \( p_a = 0.71 \) as in the benchmark.\textsuperscript{39} In Column 4 in Table 6, the switch from

\textsuperscript{37}I remain agnostic with respect to the means used to eliminate evasion. Hindriks et al. (1999) suggest that tax officials could be paid commissions on high income reports.

\textsuperscript{38}This subsidy is just subtracted from the price of the final good, \( P \). From the vantage point of firms, this acts as a subsidy on production.

\textsuperscript{39}This certainly reduces costs in terms of the number of firms to be handled since small firms are now left out of the auditing process. But such a switch would require a costly improvement in administrative technology to
the effective to the official VAT threshold generates a reallocation of 23% of small firms toward the medium (33%) and large (10%) size bins. As expected, the missing middle in terms of number of employees is alleviated. However, the gain in aggregate labor completely offsets the gain in aggregate productivity.

In Column 5 in Table 6, the wage increases by 7% to keep aggregate labor as in the benchmark. Hence most of the reallocation of labor occurs in the medium size range. The increase in productivity per worker is 7%.

In Column 6 in Table 6, I keep aggregate labor constant and introduce a subsidy on consumption to keep taxes collected as in the benchmark. In this case, consumption is subsidized with $\tau_c = -.03$. The wage increases by 13%. Overall, there is a non-negligible 16% increase in productivity per worker relative to the benchmark. Hence, a mismatch between effective and official enforcement has important implications for cross-country variation in productivity.

4.3.4 Financial Frictions

One restrictive assumption in the model is that all entrepreneurs finance their investment with internal funds. In this section, I remove this assumption to check the robustness of the main finding in this paper. Following Buera et al. (2009), I introduce capital markets in the model. Entrepreneurs can now access external financing. They can rent capital after observing their productivity level. The amount of capital that can be rented depends on the collateral an entrepreneur has to offer. Profits when an entrepreneur does not default are:

$$PF(z, k, l) - wl + (1 - \delta)k^o - rk^r$$ (29)

where $k$ is total capital, $k^o$ is capital financed internally and owned by the firm, $k^r$ is capital rented and $r$ is rate of rental of capital.

When an entrepreneur defaults, he does not post any payment to rented capital and keeps $k^r$. However, depending on the level of financial development, a portion $\phi$ of the total productive capital held by the defaulting firm, ($k^o + k^r$) is recovered by the judicial. If the entrepreneur defaults, profits are:

$$PF(z, k, l) - wl + (1 - \phi)(1 - \delta)(k^o + k^r)$$ (30)

I examine equilibria where an incentive compatibility constraint is such that entrepreneurs are indifferent between defaulting or not. Therefore the maximum amount of capital that can be rented by an entrepreneur with collateral $k^o$ is given by:

track firms and compile their information. I do not have data on administrative costs for the VAT in Uganda. I leave the evaluation of this tradeoff for future research.
\[ k^r \leq \frac{\phi(1 - \delta)}{1 + r - \delta - \phi(1 - \delta)} k^o \]  

(31)

The benchmark economy is calibrated based on \( \phi = 0 \). When \( \phi > 0 \), I need to introduce a rental rate of capital in the model. I rely on the IMF estimate for the commercial banking lending rate in 1997 in Uganda which is \( r = 0.21 \). The following results do not depend on the level of the rental rate.\textsuperscript{40}

In Table 7, I examine the effect of lowering the official VAT threshold from $50,000 to its pre-1995 level of $20,000. I look at two scenarios: 1) when markets are missing (the benchmark, i.e., \( \phi = 0 \)) and 2) when credit markets work perfectly (\( \phi = 1 \)). Aggregate labor is kept constant in all experiments of Table 7. Column 1 exhibits the effect of moving away from the benchmark to $20,000 in the absence of markets for capital. Aggregate production increases by 7%. Since aggregate labor is adjusted to be as in the benchmark, this clearly represents an improvement in productivity per worker. When entrepreneurs choose one of their inputs (capital) before the realisation of productivity shocks, a reduction in the official VAT threshold increases in efficiency. The intuition is as follows. Large firms undertake most of the production (83%). A reduction of the official VAT threshold reduces the aggregate effect on the intertemporal choice of capital because there are fewer large firms affected by the distortion.

<table>
<thead>
<tr>
<th>Table 7: Capital Market Experiments</th>
<th>(% from the benchmark)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>7 8 9</td>
</tr>
<tr>
<td>Threshold</td>
<td>No market  Perfect market</td>
</tr>
<tr>
<td>( \Delta \text{wage} )</td>
<td>20,000$  50,000$</td>
</tr>
<tr>
<td>( \Delta \text{ShareSmall} )</td>
<td>-11 -48 -48</td>
</tr>
<tr>
<td>( \Delta \text{ShareMedium} )</td>
<td>23 65 65</td>
</tr>
<tr>
<td>( \Delta \text{ShareLarge} )</td>
<td>-1 24 24</td>
</tr>
<tr>
<td>( \Delta \text{Agg.Capital} )</td>
<td>6 36 36</td>
</tr>
<tr>
<td>( \Delta \text{Agg.Output} )</td>
<td>7 20 20</td>
</tr>
<tr>
<td>( \Delta \text{Agg.Taxes} )</td>
<td>7 20 20</td>
</tr>
<tr>
<td>( \Delta \text{Agg.Bribe} )</td>
<td>9 22 22</td>
</tr>
</tbody>
</table>

In Column 2 and 3 of Table 7, I present a situation where financial frictions are removed. Columns 2 and 3 presents percentage changes from the benchmark when the official VAT thresh-

\textsuperscript{40}Goyette and Gallipoli (unpublished results) examine the efficiency implication of financial frictions. Here, I am only interested in the interaction of financial frictions and the level of the VAT threshold.
old is $20,000 and $50,000 respectively. It is clear in this case that changing the official VAT threshold generates no gain in productivity per worker when markets for capital work adequately. The introduction of perfect market for capital removes the distortion created by the interaction of the VAT threshold and financial frictions. These conclusions are not affected by using a different rate of rental $r$ or reducing the effective VAT threshold in terms of the number of employees.

The last set of experiments have important implications for the implementation of a VAT in countries suffering from financial frictions as the choice of a threshold has an impact on aggregate efficiency. This result should be contrasted with Keen and Mintz (2004) who argue that a lower VAT threshold increases administrative costs due to the increase in the number of firms to be handled. Clearly, there is a tradeoff between administrative costs and efficiency losses.\footnote{Due to data limitations, I cannot estimate the administrative costs associated with a reduction in the VAT threshold.}

Finally, the main result of the paper is not affected by the assumption of incomplete capital markets. Indeed, I observe the same quantitative effect of a switch from effective to official VAT threshold with or without the presence of financial frictions.\footnote{Results for the situation with complete capital markets available upon request.} Moreover, the value of the rental rate as no effect of the main conclusion of this paper. However, in the case with financial frictions, the main result of the paper should be considered as an upper bound. To see this, suppose the true effective VAT threshold is actually lower than the estimate presented in this paper. There are two implications. First, the benchmark efficiency loss would be smaller. And then, the switch from effective to official VAT threshold implies a smaller gain in efficiency.

\section{Conclusion}

In this paper, I examine the efficiency losses from enforcing the VAT based on a threshold measured in number of employees rather than sales. Amending a growth model with entry and exit of heterogeneous firms to include audit and tax distortions and capital accumulation generates an equilibrium firm size distribution, and patterns in firm’s share of output and age, similar to those observed in micro-data from Uganda. The numerical counterpart of the model fits the data reasonably well in several dimensions that were not explicitly targeted.

Counterfactual experiments show that: 1) a change in bargaining power has ambiguous effects on efficiency; 2) there is a 45\% loss in efficiency due to evasion; 3) switching from the effective size threshold based on the number of employees to the official VAT threshold based on sales increases productivity per worker of 16\%; 4) decreasing the VAT threshold increases output per worker when one of the inputs into production is chosen intertemporally due to uncertainty.

One caveat in the analysis is the lack of data on Ugandan administrative costs associated...
with the VAT. Proponents of the VAT argue that one advantage of increasing the VAT threshold is that it reduces administrative costs. I show that when there are financial frictions there exists a tradeoff between efficiency and these administrative costs. The quantitative assessment of this tradeoff is left for future research.

6 Acknowledgements

I would like to thank participants at the CEA 2011 in Ottawa and the 2011 EC² Conference in Florence who made useful comments. All errors are my own.

References


equilibre gnral Application au cas camerounais. Cahier de recherche CREFA.

of Public Economics (89), 599–623.


Gang, I., Goswani, O., Sanyal, A., 1988. Taxes, Fines, and Bribes: A Model of Indian Public
Finance. Markets in Developing countries: Parallel, Fragmented and Black.

Gauthier, B., Gersovitz, M., 1997. Revenue erosion through exemption and evasion in Cameroon,

Gauthier, B., Reinikka, R., 2006. Shifting tax burdens through exemptions and evasion: an


Goyette, J., Gallipoli, G., unpublished results. Distortions, Efficiency and the Size Distribution
of Firms. Working Paper.

1185–1202.

Hall, R., Jones, C., 1999. Why Do Some Countries Produce So Much More Output Per Worker

Hayashi, F., 1982. Tobin’s marginal q and average q: A neoclassical interpretation. Econometrica:

Economics 74 (3), 395–430.

60 (5), 1127–1150.

and Control 16 (3-4), 621–653.


A Theoretical Appendix

A.1 Analytical Example

The analytical expressions for labor and capital depending on the status of a firm are:

\[ l^{us} = z \left( \left( \frac{\gamma}{r} \right)^{\gamma} \left( \frac{\eta}{w} \right)^{1-\gamma} \right)^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (32)

\[ k^{us} = z \left( \left( \frac{\gamma}{r} \right)^{1-\eta} \left( \frac{\eta}{w} \right)^{\eta} \right)^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (33)

\[ l^c = z_o \left( \left( \frac{\gamma}{r} \right)^{\gamma} \left( \frac{\eta}{w} \right)^{1-\gamma} \right)^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (34)

\[ k^c = z_o \left( \left( \frac{\gamma}{r} \right)^{1-\eta} \left( \frac{\eta}{w} \right)^{\eta} \right)^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (35)

\[ l^{ul} = z \left[ (\frac{\gamma}{r})^{\gamma} \left( \frac{\eta}{w} \right)^{1-\gamma} \right]^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (36)

\[ k^{ul} = z \left[ (\frac{\gamma}{r})^{1-\eta} \left( \frac{\eta}{w} \right)^{\eta} \right]^{\frac{1}{1-\gamma-\eta}} \]  \hspace{1cm} (37)

To obtain \( z^{ul} \), I use the following the fact that this productivity threshold is such that \( V^c = V^{ul} \). Hence,

\[ (z^{ul})^{1-\gamma-\eta} \left\{ \lambda (k^{ul})^{\gamma} (l^{ul})^{\eta} - (k^c)^{\gamma} (l^c)^{\eta} \right\} - w \left[ l^{ul} - l^c \right] - r \left[ k^{ul} - k^c \right] = \theta \psi A \]  \hspace{1cm} (38)

Although, one cannot isolate \( z^{ul} \) in an obvious way from this expression, it is easy to see that \( z^{ul} \) is a function of parameters only. Note that an increase in \( z^{ul} \) corresponds to a decrease in the size of the formal sector for a fixed measure of entrepreneurs.
A.2 Proof of Proposition 1:

I explore four cases.

**Case 1:** $\psi_A \to 0$, large elasticities

When enforcement is low, we only need to consider RHS1 and RHS2. We have,

$$\frac{dB}{d\theta} < 0 \text{ if } \theta > \frac{\gamma \varepsilon_k f_\theta + \eta \varepsilon_{1f_\theta}}{1 + \gamma \varepsilon_k f_\theta + \eta \varepsilon_{1f_\theta}} \tag{39}$$

If the elasticities are large the RHS of equation 39 goes to one and $\theta$ has to be very large in order for an increase in bargaining power to decrease the bribe. In this case, entrepreneurs could increase their allocation of capital and labor by so much in response to an increase in bargaining power that they need to pay a bigger bribe than if their bargaining power had not changed.

**Case 2:** $\psi_A \to 0$, small elasticities

In this case, Equation 39 goes to zero. If the elasticities are small, an entrepreneur does not change its allocation of capital and labor by much in response to the change in its bargaining power, but it has more power to negotiate a lower bribe.

**Case 3:** $\psi_A > 0$, large elasticities

As before, let us obtain a condition that relates the equilibrium bribe to $\theta$:

$$\frac{dB}{d\theta} < 0 \text{ if } \theta > \frac{\gamma \varepsilon_k f_\theta + \eta \varepsilon_{1f_\theta}}{1 + \gamma \varepsilon_k f_\theta + \eta \varepsilon_{1f_\theta} - \frac{\psi_A}{\tau p e^{-1 - \gamma - \eta k^\gamma l^n}}} \tag{40}$$

Again if elasticities are large, the RHS of Equation 40 goes to one, and entrepreneurs need a large bargaining power to negotiate lower bribes. Note that the extra term in the denominator makes this requirement more difficult to satisfy than in Case 1. However, if $\psi_A \geq \tau p e^{k^\gamma l^n}$, an entrepreneur prefers to repay its surplus from evasion than a bribe.43

**Case 4:** $\psi_A > 0$, small elasticities

When elasticities are small, the bribe normally decreases with an increase in bargaining power. This is because an entrepreneur does not change its allocations by much in response to the change in bargaining power.

**End of Proof of Proposition 1**

43 In such a situation, an entrepreneur pays a bribe higher than its liabilities, for any level of bargaining power.
A.3 Effect of a Change in the Official Tax Rate

An entrepreneur reacts to a change in $\tau_o$ with an inversely proportional response in the choice of his inputs as shown in the following two expressions:

$$\frac{dk^{ul}}{d\tau_o} = -\frac{1}{2}(2-\theta) \frac{1}{(1-\gamma-\eta)} \lambda^{1-\gamma-\eta} z \left[ \left(\frac{\gamma}{r}\right)^{1-\gamma} \left(\frac{\eta}{w}\right)^{\eta} \right]^{-\frac{1}{1-\gamma-\eta}} < 0 \quad (41)$$

$$\frac{dl^{ul}}{d\tau_o} = -\frac{1}{2}(2-\theta) \frac{1}{(1-\gamma-\eta)} \lambda^{1-\gamma-\eta} z \left[ \left(\frac{\gamma}{r}\right)^{\gamma} \left(\frac{\eta}{w}\right)^{1-\gamma} \right]^{-\frac{1}{1-\gamma-\eta}} < 0 \quad (42)$$

Taking a derivative of $B(z,k,l)$ with respect to $\tau_o$, I get:

$$\frac{dB}{d\tau_o} = \frac{(1-\theta)}{2} z^{1-\gamma-\eta} (k^{ul})^\gamma (l^{ul})^\eta \left[ 1 + \gamma \varepsilon_{k^{ul}\tau} + \eta \varepsilon_{l^{ul}\tau} \right] \geq 0 \quad (43)$$

where

$$\varepsilon_{k^{ul}\tau} = \frac{\tau}{k^{ul}} \frac{dk^{ul}}{d\tau_o} < 0$$

and

$$\varepsilon_{l^{ul}\tau} = \frac{\tau}{l^{ul}} \frac{dl^{ul}}{d\tau_o} < 0$$

From Equation 43, note that an increase in the official tax rate has ambiguous effect on the level of the bribe. The overall effect depends on the magnitude of the elasticities of the inputs with respect to the tax rate. Suppose these elasticities are small, the level of the bribe then increases with the tax rate, that is, firms do not change much their allocation in response to a tax increase but a higher tax rate allows bureaucrats to negotiate bigger bribes. Suppose these elasticities are large, firms decrease their hiring of inputs in response to a tax increase, and this leaves less profits to be taxed from, and in turn, lower bribes to be exchanged with bureaucrats.

Finally I examine the effect of a change in $\tau_o$ on the shock-threshold $z^{ul}$. Taking the derivative of Equation 38 with respect to $\tau_o$ and using the equilibrium conditions for wages and the rental rate, I obtain:

$$\frac{dz^{ul}}{d\tau_o} = \left[ \frac{1}{2}(2-\theta) \right] z^{ul} (k^{ul})^\gamma (l^{ul})^\eta \frac{1}{(1-\gamma-\eta)\Lambda} > 0 \quad (44)$$

An increase in the official tax rate has the expected effect of shrinking the shares of unconstrained large firms. Indeed, for a fixed measure of firms with upper bound at $z_{up}$, an increase in $z^{ul}$ implies that $z_{up} - z^{ul}$ decreases.
A.4 Capital and Labor Elasticities

Proposition A1: In the analytical example, $\epsilon_{kul\theta} = \epsilon_{kul\theta}$

Proof:
I use equations 36, 37, 22 and 23 to show this result.
I use Equation 37 for $k_{ul}$ in Equation 22 to show that: $\epsilon_{k_{ul}\theta} = \frac{\theta}{k^{\mu}} \frac{dk_{ul}}{d\sigma} = \frac{\theta r_o}{(1-\gamma-\eta)\lambda}$. The exact same result obtains using 36 and 23.

End of proof of proposition A1.
B Numerical Appendix

B.1 Obtaining a Measure of Capital

There is no direct measure of capital in the Ugandan data. Calculating the quantity of capital owned by firms seems to be a daunting task for economists when there are no direct measures available. Attempts at estimating capital usually rely on perpetual inventory methods and specific, thus limiting, frameworks (Baily et al., 1981; Soderbom and Teal, 2004). Given I do not have access to firms’ initial level of capital stock in the sample, any Perpetual Inventory Method can not be implemented. Also, there is no way I can trace back the market value of the firms in our sample from the Uganda Securities Exchange given the confidential nature of the survey. This rules out Hall (2001)’s approach or other methodologies that rely on Tobin’s q (Hayashi, 1982; Hopenhayn, 1992b; Chung and Pruitt, 1994). This is also true of other methodologies used in finance to back out the value of a firm or of its capital, including the Free Cash Flow Method (Kaplan and Ruback, 1995), Capital Cash Flow method (Ruback, 2002), Adjusted Present Value method (Myers, 1974).

I suggest an alternative approach based on the information available in the dataset on interest payments, the replacement value of machinery and equipment, and the resale value of plant, land, equipment, machinery and other assets. I approximate the rate of rental of capital ($r$) using interest payments ($int$) and the resale (scrap) value of capital ($V_0$) with the following relationship:

\[
\frac{int}{V_0} = \frac{r(p_0k)}{p_0k} = r
\]  

(45)

where $k$ is scrap capital and $p_0$ its price. Note that this ratio gives me the cost of servicing existing capital. Rates of rental of capital calculated for each firms are thus likely to be lower than interest rates on outstanding debt. Furthermore, if I assume that the price of capital is the discounted sum of the returns (rate of rental minus depreciation) on a unit of consumption invested as capital, it can be expressed as:

\[
p_0 = \sum_{t=0}^{\infty} \left( \frac{r - \delta}{1 + r} \right)^t = \frac{1 + r}{1 + \delta}
\]  

(46)

where $\delta$ is the rate of depreciation. Each firm’s own price of capital can be used to identify its own stock of capital with the following relationship:

\[
\ln(k) = \ln(V_0) - \ln(p_0)
\]  

(47)

\footnote{This is by no means an exhaustive list of the subject.}

44