

Distortions, Efficiency and the Size Distribution of Firms*

Giovanni Gallipoli
University of British Columbia

Jonathan Goyette
Université de Sherbrooke

Abstract

Microdata information about firms' input choices and effective tax liabilities is used to quantify the extent of resource misallocation and efficiency losses due to large tax distortions and limited access to credit. We develop an equilibrium model of firms' behavior in which the tax and credit environments act as a selection mechanism restricting the growth of all but the most productive firms. We show that such a model, parameterized and validated using a variety of data restrictions, has the potential to rationalize several puzzling observations about firms' input choices, size and growth patterns. Counterfactual experiments are designed to gauge the losses associated to different deviations from first-best. We find that firms' optimal responses to the tax distortions are quite effective in reducing efficiency losses. As a consequence, tax distortions only account for a small proportion of the gap between an undistorted first best economy and the benchmark. On the other hand limited and costly credit is associated to more significant misallocation of productive resources and leads to much larger aggregate efficiency losses, accounting for at least 90% of the difference in output per worker between first-best and benchmark economy.

*We thank participants at various conferences and seminars for helpful discussions. Paul Beaudry, Patrick Francois, Dirk Krueger, Henry Siu, Vadim Marmer and Makoto Nakajima provided insightful comments. All errors are ours.

1 Introduction

Cross-country differences in output levels and growth rates are both large and persistent (Hall and Jones, 1999; Klenow and Rodriguez-Clare, 1997). Understanding the determinants of such discrepancies is especially crucial for less developed countries (LDCs) suffering from high poverty rates. As pointed out by Prescott (1998) and Easterly and Levine (2001) much of the measured variation is attributable to heterogeneity in total factor productivity (TFP) and much recent research has attempted to develop empirically sound theories of TFP. One view is that institutional distortions have large effects on the allocation of productive resources and, therefore, on measured productivity (see North et al., 1981; Acemoglu et al., 2001; Restuccia and Rogerson, 2008; Alfaro et al., 2007). However, we only have a limited understanding of the mechanisms mapping specific distortions into observed productive choices and ensuing efficiency losses. For this reason Restuccia and Rogerson (2008) emphasize the necessity to obtain better measures of specific distortions and to evaluate their aggregate consequences.¹ This paper documents a large and easily measurable input market distortion in Uganda (see also Goyette, 2009) and exploits this information to back out the optimal input choice responses of firms. Building on original work by Hopenhayn (1992) a model of dynamic firm's choice is developed in an environment characterized by suboptimal tax regulation and endogenous credit constraints. As in a standard model, firms optimally choose their capital-labor mix conditional on current and expected idiosyncratic productivity. Moreover, two simple extensions introduce path-dependence in input choices. First, we assume that productive capital can be purchased one period ahead, before observing the productivity shock, or rented at market rate in the same period production takes place, after observing the shock. However, access to the capital rental market depends on the amount of collateral available, consisting of own accumulated capital. This market arrangement allows us to characterize the degree of credit markets' completeness. When credit markets work well, all firms are able to borrow the ex-post optimal amount of capital, after observing their productivity draw. Viceversa, when credit market frictions exist, access to rental markets is limited, meaning that smaller firms may be shut out from borrowing. The second twist on the standard model is the introduction of size-dependent tax liabilities: empirical evidence suggests that uneven tax auditing results in firms with more employees paying higher taxes on their revenues; in fact, our analysis suggests that tax liabilities jump in a discrete fashion when the number of employees grows above a certain threshold. The interaction of limited credit market access and

¹Along these lines, a new applied literature is flourishing. Hsieh and Klenow (2007) show that distortions putting a wedge between the marginal product of capital and labor may reduce efficiency by 30-50% in China and 40-60% in India. Guner et al. (2008) and Garcia-Santana and Pijoan-Mas (2010) examine the costs of size-dependent policies distorting production scale. Buera and Shin (2008), Midrigan and Xu (2009) and Buera et al. (2009) examine the effect of financing constraints on TFP.

size-dependent tax liabilities leads to endogenous dispersion in capital-labor ratios, revenues and profits for firms with similar idiosyncratic productivity.

The model is parameterized in two steps: first some parameters are assigned using external estimates; then, the model is simulated in order to replicate other data features. The parameterized model is benchmarked using a variety of data restrictions and then used to perform some counterfactual exercises.

The parameters for the fiscal environment are calibrated using firm-level information from Ugandan data. We use technology parameters from the existing literature. Finally, we estimate some parameters using the model.²

The model can, to a large extent, reproduce peculiar features of the data that were not explicitly targeted. In particular, as observed in the firm-level Ugandan data, the model generates a clustering of small firms right before the sharp increase in audit intensity. These small firms choose to scale up production by substituting capital for labor, as their productivity shocks are not large enough to offset any tax losses associated to ‘formality’. We note that these choices generate a gap in the size distribution of firms (as measured by the number of employees) and result in the so-called ‘missing middle’ phenomenon, typical of so many LDCs.³ We examine the theoretical implications of the model by sequentially removing different distortions. We show that the increase in audit intensity is key in generating the peculiar input choices observed in the data for small firms. However, removing all tax distortions would close the gap in output per worker between the benchmark and a first-best undistorted economy only by 5%. Removing external financing constraints and reducing the cost of credit generates much larger gains and explains roughly 95% of the output difference between our benchmark economy and the first-best outcome. Finally, we find that the ‘missing middle’ in the firms’ size distribution is a by-product of the audit distortion in the tax system and is associated to negligibly small losses in aggregate efficiency. Our results suggest that the ‘missing middle’ simply reflects the optimal responses of forward looking firms to the tax and auditing environment, and that such responses are quite effective in reallocating resources so to minimize the impact of the distortion.

1.1 Related Literature

Fiscal Environment It is often argued that tax administrations in less developed countries (LDCs) target larger firms more intensively because tax collection is expected to be greater than

²Either by directly setting them to match specific data features, or through a ‘distance minimization’ approach (Gourieroux et al., 1993; Smith Jr, 1993).

³Many papers in the development literature argue that LDCs exhibit too few medium firms relative to developed economies like the US (Tybout, 2000; Liedholm and Mead, 1987; Little et al., 1987; Steel, 1993; Steel and Webster, 1992; Sleuwaegen and Goedhuys, 2002).

enforcement costs (de Soto,1989; Gauthier and Gersovitz,1997; Gauthier and Reinikka,2006). Goyette (2009) shows that the probability of being audited by tax officials in Uganda rises sharply around a certain threshold corresponding to about 30 employees. This translates in a significant ‘break’ in the density distribution of firm sizes around that threshold. Data suggest that entrepreneurs, who would otherwise grow their firms beyond that threshold, choose suboptimal capital-labor allocations in order to avoid steep increases in regulation costs. In response to a positive productivity shock a large proportion of entrepreneurs choose to scale up production by substituting capital for labor. This results in the observed clustering of many firms at or below the 30-employee threshold, and into significantly higher capital-labor ratios and significantly lower growth rates for smaller firms. Only relatively large productivity shocks outweigh the costs of increasing employment beyond the ‘formality’ threshold.

Financial constraints Recent work has highlighted how credit market imperfections and heterogeneity in access to external financing may distort the allocation of resources in developing countries and result in substantial productivity losses. Banerjee and Duflo (2008) use variation in access to a lending program to evaluate the extent of credit constraints faced by Indian firms. They argue that constrained firms use credit to expand production while unconstrained firms use credit as a substitute for other borrowing. Their results suggest that many firms may be credit constrained and have very high marginal rates of return to capital.⁴

Buera and Shin (2008) quantitatively evaluate the impact of financial frictions, and resource misallocation, on the dynamics of economic development. They show that, given some misallocation of initial resources, financial frictions delay the necessary adjustment. According to the authors, financial frictions are responsible for between a quarter and a half of total misallocation in LDCS, a result in line with Restuccia and Rogerson (2008). Arellano et al. (2008) examine how financial development influences firms’ financing and growth and show that high credit costs limit debt disproportionately for small firms thus making their scale inefficient. Buera et al. (2009) provide evidence that poor countries are less productive in the tradable and investment goods’ sectors. They develop a quantitative framework to explain these cross-country patterns and, as in Restuccia and Rogerson (2008), they find that variation in financial development can explain a two-fold difference in output per worker across countries. Interestingly, Midrigan and Xu (2009) find that aggregate TFP losses from misallocation of factors are not as large as these previous studies have found using a parametrization that focuses on the returns to capital and other factors rather than the size of the financial sector.

The rest of the paper is structured as follows. Section 2 presents empirical evidence which

⁴However, Banerjee and Duflo (2005) explain that even if numerous small firms are credit constrained their share of total capital is too small to explain cross-country differences in output per worker. Incidentally, our results partially confirm this finding, as most of the production is carried out by the largest firms in equilibrium.

motivates this paper. Based on these facts, we develop a dynamic model with heterogeneous firms in section 3. Section 4 summarizes the calibration strategy and the numerical implementation of the model and discusses the results of counterfactual experiments. Section 5 concludes.

2 Some Facts

This section describes some relevant features of a cross-section of Ugandan firms. Data are sourced from the Ugandan Enterprise Survey (UES) initiated in 1998 by the World Bank and by the Uganda Private Sector Foundation. Firms were randomly selected and are a representative sample of Ugandan firms. 243 firms were interviewed about their activities between 1995 and 1997. The sample covers businesses from five economic sectors: commercial agriculture, agro-processing, manufacturing, tourism and construction, and five geographical areas.⁵ The survey focuses on firms' activities including investment, sales, finance, regulation, infrastructure, taxation, corruption and labor market.

In what follows we classify firms in 3 size groups: small firms have 30 employees or less, medium firms have between 31 and 75 employees, and large firms have 76 employees or more. This categorization is consistent with our own findings about regulation costs, as well as with previous literature; for example, Tybout (2000), Gauthier and Reinikka (2006), Soderbom and Teal (2004), Goyette (2009) and many other papers in the development literature use a similar classification of firm size.⁶

Before providing some details about the empirical analysis, we list some noteworthy features of the Ugandan data: (1) audit intensity of smaller firms, as reported by entrepreneurs, depends only on the number of employees, but not on capital or revenues; (2) the audit intensity increases sharply for firms with more than 30 employees; (3) small firms finance most of their investment with their own cash flow, while larger firms have access to credit; (4) the size distribution of firms exhibits a missing middle (too few medium firms), a common observation in LDCs; (5) there is a significant mass of small firms before the threshold at which audit intensity sharply increases; (6) small firms exhibit lower growth rates (of revenues) than their larger counterparts; (7) small firms exhibit higher capital-labor ratios than their larger counterparts.

Audit intensity Table 1 reports OLS estimates from a regression of the audit intensity on firm size and other potential determinants of auditing, namely capital, sales and tax obligations. The sample is divided in two groups, small and large. The first group includes firms with 30

⁵Namely, Kampala, Jinja-Iganga, Mbale-Tororo, Mukono and Mbarara. For a detailed discussion about the data and summary statistics see Goyette (2009).

⁶Results still hold if we use a classification where small firms have 20 employees or less, medium firms have between 21 and 100 employees, and large firms have 101 employees or more.

employees or less, the second group consists of firms with more than 30 employees. For small firms the number of employees is the only predictor of audit intensity. For larger firms, capital and sales are also significant covariates.

Column 1 through 3 of Table 1 show that firm size (as measured by the number of employees) is a significant predictor of auditing for small firms. For larger firms the coefficients on sales, or taxes, are also positive and significant. Capital stock is never significant. This seems to confirm the finding, in the development literature, that auditing of small firms has some discretionary nature: if tax auditors have to roam around town to physically identify small firms, it is common practice for small firms in LDCS to pass under the authority’s radar (de Soto,1989; Gauthier and Gersovitz,1997; Gauthier and Reinikka,2006).

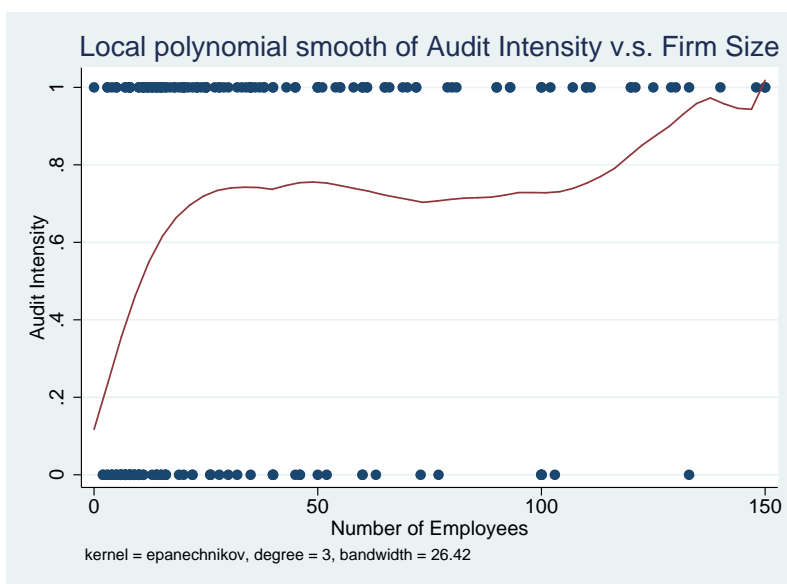


Figure 1: Audit Intensity v.s. Firm Size in Uganda in 1997

Moreover, graphical evidence (see figure 1) also suggests that that entrepreneurs with small firms are audited less intensively than their larger counterparts. In fact, there seems to be a step increase in the auditing probability around a certain size.⁷

In our model we approximate the pattern observed in figure 1 with the simplest polynomial form, a step function. We do this by using the firm size category defined above. In figure ?? the first step corresponds to the average audit intensity faced by small firms, equal to 0.53; the second step and third steps represent the average audit intensity faced by, respectively, medium firms (0.71) and large firms (0.77). A simple t-test of the difference in average audit probability between small and medium firms is significant (see Goyette,2009).

⁷The figure plots a local smoother with an Epanechnikov kernel, a polynomial of degree 3 and a bandwidth of 26.5 employees.

Table 1: Determinants of Audit intensity

		Dependent Variable: Audit intensity		
Explanatory variable:		1	2	3
# of Employees	small ^e	.26***	.24***	.20**
		(.12)	(.07)	(.06)
	large ^e	-.002	-.05	.07**
		(.06)	(.05)	(.03)
Capital	small	-.01		
		(.30)		
	large	.03		
		(.03)		
Sales	small ^e		.02	
			(.02)	
	large ^e		.07***	
			(.02)	
Taxes	small ^e			.02
				(.02)
	large ^e			.07***
				(.02)
R^2		.18	.23	.22
F1 ^c		3.22**	9.75***	8.91***
p-value		[.02]	[0.0]	[0.0]
F2 ^d		1.84	3.45***	2.30*
p-value		[.14]	[0.02]	[0.08]
Nb. Obs.		150	225	224

^a std. dev. in parenthesis; number of firms in curly brackets; t-statistics in brackets; significance: *=10%, **=5%, ***=1%. ^b All regressors in logs; all regressions include dummies for industry-location.

^c F1: Wald test statistic. H_0 =coefficients jointly zero.

^d F2: Wald test statistic. H_0 : coefficients equal. ^e Small firms: 30 or less employees. Larger firms: all others.

Cash flow constraint In our sample 53% of firms (mainly small firms and a few medium ones) reported being constrained (that is, unable to request a loan) between 1995 and 1997. The main reported reason was that interest rates or the collateral requirement were too high. As a consequence, small Ugandan firms have to finance investment with their own cash flow.

Unfortunately our data does not provide information about the outstanding debt level of firms, or the ratio of equity to debt. However, the dataset gives information about the main sources of financing. Table 2 confirms the limited access to credit for small firms by showing the correlation between different sources of financing and firm size.⁸

Table 2: Correlation Coefficients between Sources of Financing and Firm Size

Source of financing	Correlation
a. Own profits	-.12 (.12)
b. Personal savings	-.25*** (.001)
c. commercial banks	.11 (.16)
d. development finance	.24*** (.003)
e. parent or holding company	.19** (.02)

Significance: *=10%, **=5%, ***=1%; p-value in parenthesis; Other sources of financing are family and relatives, suppliers credit, money lender, sale of assets, lease finance, new partner; all other sources yield coefficients below 0.1 and are positively related to size. Number of observations is 157 for a. and b., and 156 for c., d. and e.

Table 2 shows a significant and negative correlation between firm size and financing through own personal savings. Larger firms have easier access to development finance and loans from parent or holding companies.

So far, we have documented some facts about tax and credit distortions. In what follows we discuss three peculiar features of the data which, we argue, might be related to those distortions: (1) the Ugandan firms' size distribution exhibits too few medium firms, relative to a developed economy; (2) revenue growth rates of small firms are lower than those of their larger counterparts; (3) observed capital-labor ratios in small firms are higher than those of their larger counterparts.

⁸Sources are: profits, personal savings, family, relatives and friends, commercial banks, development finance, supplier's credit, money lender, parent or holding company, sale of assets, lease finance, or new partner.

Density of firms’ sizes Figure 2 plots the density of firms’ sizes (estimated with size bins of 5 employees and a bandwidth of 50 employees). The density is estimated using the McCrary (2008) local density estimator.⁹ This estimator allows us to test for a discrete change in the density of firms’ sizes around the threshold of ‘formality’, that is, in the area where effective tax liabilities suddenly increase. A discontinuity in the density can be clearly identified at 30 employees (vertical line in Figure 2): The difference in the intercepts at that point is equal to $-.99$ and its standard deviation is 0.3 ($p\text{-value}=0.0$). This evidence suggests that entrepreneurs’ labor demand choices might be distorted around the size threshold of 30 employees. One could conjecture that the sharp increase in audit intensity around 30 employees – documented above – might be key in explaining this discontinuity. However the discontinuity could also be the result of financial frictions: interest rates and collateral requirements are such that only the most productive firms might be able to grow above the 30 employees threshold. Additional data evidence and numerical simulations presented in the rest of the paper suggest that the step increase in audit intensity is mostly responsible for the shape of the firms’ size distribution.

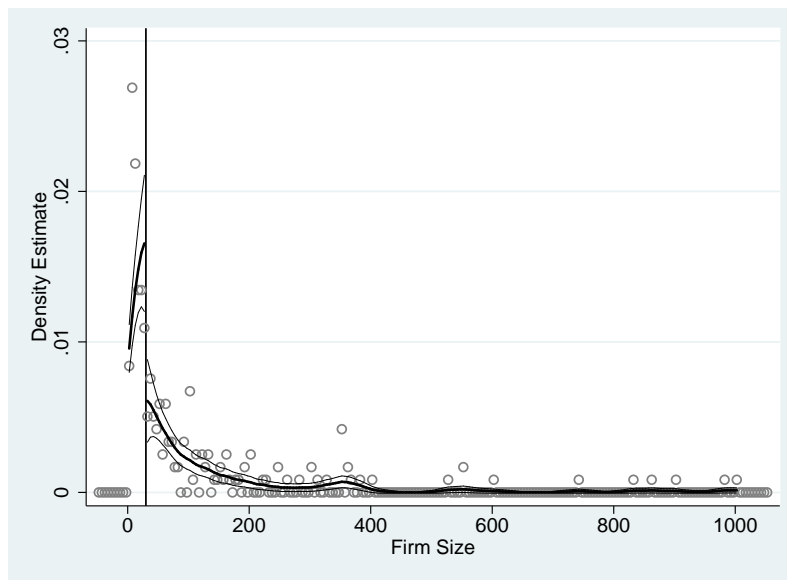


Figure 2: Density of Firms’ Sizes

⁹The estimator is an extension of the local linear density estimator. In a first step, we obtain a finely gridded histogram of the size variable (number of employees). In the second step, we smooth this histogram 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. We use a weighted regression where more weight is given to the bins nearest where we are trying to estimate the density, 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.

Why 30 employees? There are two main reasons why we chose to set a ‘formality’ threshold at 30 employees. First, based on our own analysis, the upper bound on the size bin of small firms at 30 employees seems to be appropriate. Table 3 reports a sensitivity analysis on the log difference in intercepts using alternative cutoff values, all between 10 and 35 employees. The largest and most significant log difference between intercepts occurs, in fact, at 30 employees.¹⁰ The fact that we observe a significant discontinuity from 15 to 30 employees using McCrary (2008)’s methodology indicates that there is a significant mass of firms within this range. The mode of the size distribution is at 15 employees. Moreover, this choice of threshold of 30 employees for small firms and 75 employees for medium firms is consistent with previous research using this dataset (See Gauthier and Reinikka, 2006).

Table 3: Comparison of Size-Thresholds for the Density of Firm Size

Size-threshold	Bandwidth (Nb. of Emp.)	Log Difference in intercepts	Standard Error of Log Difference
10	29.0	-0.13	0.24
15	32.2	-0.62	0.25
20	35.7	-0.53	0.26
25	41.2	-0.73	0.28
30	47.6	-0.97	0.31
35	51.7	-0.51	0.31

All estimates are based on size-bins of 5 employees.

Firms’ Growth Rates We compute the growth rate of revenues as:

$$FirmsGrowth = \ln(Sales1997) - \ln(Sales1996) \quad (1)$$

Average revenue growth is 0.02 for small firms, 0.13 for medium firms and 0.17 for large firms. The results of a regression of the growth rate on firm size are shown in Table 4: the coefficient on the number of employees is positive and significant. A simple calculation tells us that a 1% increase in the average number of employees (which is 20 employees in our dataset) adds 0.7% to the revenue growth rate.

¹⁰The automatic selector from McCrary (2008) suggests the use of a binsize of 22 employees and a bandwidth of 45 employees. The estimator is robust to different choices of binsize provided that $h/b > 10$. Given that a bin size of 22 employees does not convey a good graphical representation of the density around the threshold, we have used a bin size of 5 employees. The results are robust to choices of bin size between 5 and 50 employees (results not shown). The choice of bandwidth does affect the performance of the estimator.

Table 4: Revenue growth and Number of Employees: OLS Estimates

Dependent Variable: Growth rate of Revenues (1996-1997) ^a	
Number of Employees	.0003** (.0001)
Constant	.05 (.04)
R^2	.03
# of observations	208
F-test ^b	5.82**
p-value	[.02]

^a significance:*=10%,**=5%,***=1%; Std.Dev. in parenthesis, p-value in brackets;

^b F-test with H0: coefficient of the number of employees is zero.

This evidence on firms' growth suggests two possibilities: (1) small firms may choose to restrict their employee base (and revenues) to avoid being targeted by tax officials; and/or (2) financial frictions may result in suboptimal access to capital and lower growth for small firms with little collateral. We use the numerical counterpart of our model to assess the relative importance of these two distortions.¹¹

Capital-Labor Ratios A very peculiar, and informative, feature of the Ugandan data is that smaller firms appear to exhibit higher capital-labor ratios than medium firms. Figure 3 reports the correlation between capital-labor ratios and firm size for all firms between 1 and 150 employees.¹² The average capital-labor ratio for the first size group (blue dots) is larger than for medium firms (green dots). Capital-labor ratios appear to be increasing before the formality threshold, that is, just before an increase in the expected audit intensity. In the rest of the paper we link this local 'build-up' of capital-labor ratios to the optimal input choices of forward-looking entrepreneurs, conditional on effective tax liabilities and collateral constraints.

¹¹The growth patterns observed in Ugandan data contrast with some findings for the US. For example Evans (1987) and Hall (1987) argue that small firms usually exhibit higher growth rates than larger firms.

¹²This range makes the plot more compact. The result holds if we use larger firms.

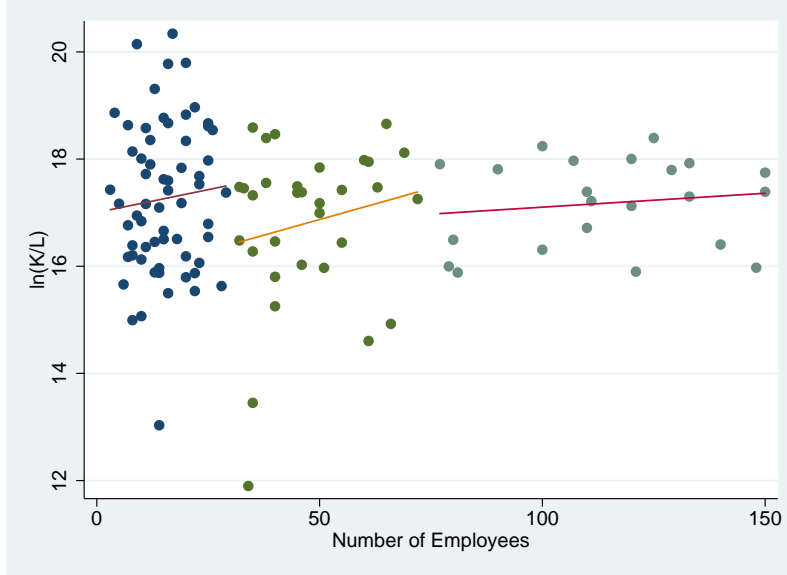


Figure 3: Capital-Labor Ratios v.s. Firm Size

3 A Model of Firm's Growth

3.1 The benchmark economy

We consider an economy consisting of entrepreneurs, workers, and government.¹³

3.1.1 Entrepreneurs

In equilibrium there is a measure 1 of entrepreneurs. Entrepreneurs are risk neutral and set up a firm to produce a homogeneous consumption good. We do not model the entry decision and assume that, in each period, new entrants fully replace exiting firms. Entrepreneurs make production decisions – that is, capital rental and labor hiring – based on: (1) their current productivity shock; (2) the stock of owned capital carried over from the previous period; (3) the likelihood of being audited. After production takes place, entrepreneurs choose how much to invest in ‘owned’ capital k^o , based on their expectations over future productivity and access to the rental market for capital.

In the benchmark economy we assume that access to the rental market for productive capital may be limited by conditions typical of a LDC, e.g. limited enforceability of contracts (the model is also consistent with constraints due to partial unobservability of revenues or incomplete contingent contracts). As a consequence, entrepreneurs with small firms must finance part or all

¹³We assume that economic agents are ex-ante different and we do not study their choices of activity, e.g. worker vs entrepreneur.

of their investment through their own cash flow and invest in their own firm through outright purchases of capital, while larger firms have generally better access to credit.

Ugandan data suggests that, on average, most entrepreneurs evade a significant share of their tax liabilities in each year. We therefore assume that, in each period, they only pay a share of their tax dues. However, in the event of being audited, they are forced to repay the whole surplus from tax evasion.¹⁴ Entrepreneurs with low capital stock and productivity may choose to exit the market at the end of the period. Every period a mass of firms exit and is replaced by a mass of new entrants in the next period.

Entrepreneurs are heterogeneous in two dimensions. Upon entry in the market each entrepreneur is endowed with a productivity parameter z_t drawn from an initial distribution function $\nu(k, z)$. Productivity shocks then evolve according to a transition function denoted as $g(z_{t+1}, z_t)$. Moreover, stocks of owned capital $k_t^o \in (0, k^{up}]$ differ across entrepreneurs because of their previous investment history.

The price of output is normalized to one, labor is remunerated at market wage w_t , while capital is rented at the rate r_t . The production function $F(z, k, l)$ takes as input capital (both owned and rented) and labor (number of employees), and it exhibits decreasing returns to scale in both inputs. We assume that:

$$F(z, k, l) = zk^\gamma l^\eta \quad (2)$$

where $\gamma + \eta < 1$. This choice of functional form allows to easily pin down the size of a firm.

Fiscal environment. We model the fiscal environment by introducing a total regulation cost $RC(l_t)$ which varies with firm's size:

$$RC(l_t) = p_a(l_t)T_t^e + T_t^p \quad (3)$$

where $p_a(l_t)$ is the probability of being audited, which varies with firm size, so that:

$$p_a = \left\{ \begin{array}{l} p_a^s \text{ if } \textit{small} \\ p_a^l \text{ if } \textit{medium/large} \end{array} \right\} \quad (4)$$

and $p_a^s < p_a^l$. An audited entrepreneur has to repay the amount of taxes evaded T^e .¹⁵ T^p is the amount of taxes actually paid by an entrepreneur. Note that $T^e + T^p = T^0$ where T^0

¹⁴The model can easily include a fine on top of the surplus of evasion, as well as bribing of tax officials, without substantial changes. Data show that audited firms only pay back a portion of total liabilities, after negotiating with bureaucrats.

¹⁵This repayment might, in part, go to bureaucrats rather than the government itself. Based on the IMF Staff Country Report for Uganda in 1998 we know that, for the period going from July to December 1997, only 16% of the assessments were refunded through ordinary tax audits. The report also states that compliance to fill tax reports was low. We thus set the fine paid by entrepreneurs to zero in equation 3.

is the official tax liability. Taxes are proportional to revenues so that $T^p = \tau_p F(z, k, l)$ and $T^e = (\tau_0 - \tau_p) F(z, k, l)$ where $F(z, k, l)$ are revenues, while τ_0 and τ_p are official and effective tax rates, respectively.

Investment financing and credit market. Entrepreneurs can finance purchases of productive capital with their own cash-flow, meaning that at the end of each period they decide how much capital to carry over for production next period. This decision is based on an expectation of future productivity. We also assume that entrepreneurs may access external financing to rent capital after observing their productivity level. Of course this lends more flexibility in input choices, as capital levels can be adjusted depending on the productivity realization. However the amount of capital that can be rented depends on the collateral that can be put forward.

Denote capital owned by an entrepreneur as k^o , and capital rented as k^r .¹⁶ The total level of productive capital in period t is $k = k^o + k^r$.

The level of financial development depends on the ability to enforce contractual obligations. We define $\phi \in [0, \phi^{max}]$ as the degree of contract enforceability, where $\phi = 0$ implies no enforceability and no access to credit markets for entrepreneurs, while $\phi = \phi^{max}$ corresponds to full enforceability of financial contracts, meaning that lenders are able to fully recover their loan and lay claim on the collateral. This flexible specification can describe different economies, see for example the discussion in Buera et al. (2009).

The level of capital that can be rented in a country with given ϕ depends on an incentive compatibility condition that equates the benefits to default to those of not doing so. We consider only equilibria where the rental contracts are incentive compatible and, thus, fulfilled. We make the assumption that entrepreneurs decide the amount of investment for the next period before making the decision to default or not. Profits in period t , if an entrepreneur does not default, are:¹⁷

$$F(z, k, l) - wl + (1 - \delta)k^o - rk^r \tag{5}$$

where $F(z, k, l)$ is the production function and δ the depreciation rate of capital. Hence, equation 5 says that profits are given by production plus the residual value of owned capital, minus payments to other production factors, i.e., capital rented and labor. If the entrepreneur defaults, profits are:

$$F(z, k, l) - wl + (1 - \phi)(1 - \delta)(k^o + k^r) \tag{6}$$

¹⁶For simplicity we do not model the supply of rental capital and assume that its price is taken as given in Uganda

¹⁷For simplicity we abstract from regulation and fixed costs, which would cancel out anyway, while solving for the condition on rented capital.

When an entrepreneur defaults, she does not post any payment to rented capital and keeps k^r . However, depending on the level of financial development, the judicial can recoup a portion ϕ of the total productive capital held by the defaulting firm, $(k^o + k^r)$. External credit markets are only willing to rent capital up to the point in which 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:

$$k^r \leq \frac{\phi(1 - \delta)}{1 + r - \delta - \phi(1 - \delta)} k^o \quad (7)$$

The entrepreneur's objective function. At the beginning of a period an entrepreneur observes current productivity and owned capital stock, then chooses the amount of labor and capital to rent so that production takes place and wages and interest income are paid to rented factors. Some entrepreneurs are audited at this point. Before deciding whether to exit the market or not, an entrepreneur chooses how to divide profits, π_t , between own consumption, c_t , and capital investments in the firm, k_{t+1}^o .¹⁸ The objective function of a risk neutral entrepreneur is:

$$\max_{c_t, k_{t+1}^o} E_0 \sum_{t=0}^{\infty} (\varphi\beta)^t c_t \quad (8)$$

subject to

$$c_t + k_{t+1}^o = \pi_t \quad (9)$$

$$c_t \geq 0$$

$$k_{t+1}^o > 0$$

$$k_0 \text{ given}$$

where φ is an exogenous probability of survival, β is the discount rate and profits are given by:¹⁹

$$\pi_t = F(z_t, k_t, l_t) - w_t l_t + (1 - \delta)(k_t^o + k_t^r) - (1 + r_t - \delta)k_t^r - RC(l_t) \quad (10)$$

¹⁸In this model an entrepreneur can be interpreted as a firm's management, deciding how much of the profits to reinvest and how much to redistribute as dividends to stakeholders.

¹⁹A fixed cost is often an important feature of this class of models à la Hopenhayn. 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 we already have two selection mechanisms (tax distortion and imperfect capital markets), it seems reasonable to set the fixed cost to zero and focus on the effect of the mechanisms of interest. Moreover, we have no direct empirical evidence to calibrate the fixed cost based using our data. Finally, calibrating fixed costs would yield one more degree of freedom to fit model to data. We prefer to impose greater discipline on the model and use a smaller set of structural parameters for the calibration.

where k^r is subject to the constraint defined in equation 7. Given current shock z_t and current level of owned productive capital, k_t^o , an entrepreneur chooses how much to produce by choosing whether to rent additional capital k_t^r and the number of employees l_t . The latter choice depends in part on whether the number of employees results in a higher audit probability. Three outcomes are possible: if the ‘unconstrained’ number of employees is below the threshold for formality (denoted as l^{ic} and equal to 30), then the firm is ‘informal unconstrained’ (or “iu”) and the audit distortion has no bite; if, however, the ‘unconstrained’ number of employees is above the l^{ic} threshold, then the firm has to consider whether it would be more profitable, given the higher chance to be audited, to set $l_t = l^{ic}$ and increase output by renting more capital. This of course depends on the current productivity shock as well as on the amount of collateral available to the firm. A firm choosing to cap its number of employees at l^{ic} and, if feasible, to boost its capital-labor ratio is denoted as an ‘informal constrained’ (“ic”) firm, because the audit distortion is binding and the choice of employees is constrained at the threshold of formality. Finally, if the firm owns enough collateral, and if the productivity shock is large enough, the firm might choose $l_t > l^{ic}$, in which case we denote it as a ‘formal unconstrained’ firm (“f”).

In summary, the interaction of the audit distortion and the collateral requirement for renting capital results in a distribution of firms’ sizes and capital output ratios which is different from what would emerge in an undistorted market. A firm can be in each of three possible states s , with $s = \{iu, ic, f\}$ which correspond to whether the audit distortion has any current bite on hiring choices.²⁰

3.1.2 Workers

There is a measure of workers sufficient to satisfy labor demand at the equilibrium wage in the benchmark economy. Each worker is risk-averse with preferences given by:

$$\sum_{t=1}^{\infty} \beta^t [u(c_t) - d(n_t)] \tag{11}$$

subject to

$$c_t \leq w_t n_t \tag{12}$$

where c_t is consumption and n_t is individual labor supply, which is either zero or one. Following Hopenhayn and Rogerson (1993), we assume that workers choose employment lotteries and have access to markets to diversify idiosyncratic risks. In the context of an LDC this amounts

²⁰Given that the difference in audit intensity between medium and large firms is not large, we do not introduce an exogenous size-threshold at the upper bound of the medium size bin. Adding this change would be easy and like improve the model fit to data. We choose to keep the model simple and assume an identical audit intensity for medium and large firms.

to assume some degree of cross-insurance among workers, possibly through the use of extended family ties or some informal credit. Workers can then be jointly represented as one representative agent with preferences given by:

$$\sum_{t=1}^{\infty} \beta^t [u(c_t) - DN_t] \quad (13)$$

where N_t is the fraction of employed workers and D is a disutility parameter. We assume throughout that $u(c_t) = \log(c_t)$ and we denote the solution to this problem as N^* .

3.1.3 Government

Government expenditure G is not valued by workers and entrepreneurs and it includes tax auditing costs. All government revenues accrue from the proportional taxation of firms revenues. We use reported tax liabilities, jointly with reported tax payments, to calibrate the effective corporate tax rate.²¹

The government budget constraint is balanced in each period and is defined as:

$$G \leq T_t^{gvt} = \int [T_t^p + p_a(l_t)T_t^e] \mu^*(k^o, z) d(k \times z) \quad (14)$$

where T_t^{gvt} denotes total government revenues and $\mu^*(k^o, z)$ is the equilibrium distribution over owned capital and productivity.²²

3.2 Stationary Equilibrium

3.2.1 Recursive notation

The solution to the problem of an entrepreneur can be represented recursively. The labor hiring decision is static, and employment l_t is chosen so that the marginal product of labor equals the equilibrium wage. One can thus solve for labor in terms of capital, and write the Bellman equation of an entrepreneur as a simple choice over capital investment subject to the constraints 7, 9, 10. Given the current number of employees, and associated formality status $s = \{iu, ic, f\}$ and audit probability, one can define the conditional value functions:

$$V^s(k, z) = \max_{k'} \{c^s + \varphi \beta CV^s(k', z')\} \quad (15)$$

²¹We do not explicitly model any labor income or consumption taxes. This could be easily included, but would not alter or add to our findings. Moreover, modeling labor taxes would be hard given the ‘informality’ issues in the Ugandan context.

²²The choices of labor and capital rental are static functions of these states.

where the continuation value, CV^s , is defined as maximum expected utility over future formality choices:

$$CV^s(k', z') = \int_{z'} \max [V^{iu}, V^{ic}, V^f] dg(z', z) \quad (16)$$

At the end of each period, before making their investment decision, entrepreneurs decide whether to stay in the market or exit. Exit occurs when the continuation value CV is less or equal to zero. The measure of firms with current state (k, z) who choose to exit the market is denoted as $ex(k, z)$.

Finally we define the ‘unconditional’ value function as the upper envelope of the value functions defined for each given ‘formality’ state s . Such value function is defined as the max over the s -conditional values:

$$V(k, z) = \max_{s \in \{iu, ic, f\}} V^s(k, z) \quad (17)$$

The exit decision is made at the end of the period, after production takes place. Only when an entrepreneur exits the market there is an opportunity for an entrant to get in the next period. In this way we abstract from modeling the possibly complex discrete decision to entry and we ensure that our computational problem remains simple. One condition to get a stationary equilibrium is to equalize the mass of entrants to the mass of entrepreneurs exiting the market. Taking entry as given, the entrepreneur’s decision problem results in four decision rules, two before and two after production takes place: (i) choice of rented capital; (ii) choice of hired labor; (iii) choice of capital to carry over to next period and, finally, (iv) exit decision. We denote these optimal decision rules as k^{r*} , l^{d*} , k^{o*} and $exit^*$.

3.2.2 Definition of equilibrium

A concise way to summarize heterogeneity is to use a probability measure defined on subsets of the individual state space (see for example Lucas and Prescott, 1971; Hopenhayn, 1992). For simplicity in what follows we use the symbol k for k^o , meaning that the measure μ summarizes the distribution of owned capital, which is the relevant variable for the dynamic decisions of entrepreneurs. Let $\mu(k, z)$ be a probability measure, describing the state of the industry in a given period of time,²³ and let the mass of entrants at the beginning of next period be denoted as en . Furthermore, let $G(k, z; k', z')$ be a transition function mapping current productivity states into future productivity states.

We posit that the total measure of exiting firms is the sum of those which are shut down by entrepreneurs, with measure $ex(k, z)$ for any given (k, z) , plus those that disappear exogenously

²³We define an industry as a continuum of firms which produce a homogeneous good (Hopenhayn, 1992).

with probability $(1 - \varphi)$. The total measure of exiting firms, ξ , is defined as:

$$\xi = \int [\varphi ex(k, z) + (1 - \varphi)\mu(k, z)] d(k \times z) \quad (18)$$

In equilibrium the mass of new entrants equals ξ , and its measure over the productivity domain is $\nu(z'|k = k_0)$. The evolution of the distribution of firms can be written as follows:

$$\mu'(k', z') = \int \xi \nu(z'|k = k_0) dz' + \varphi \int [1 - ex(k, z)] G(k', z'; k, z) \mu(k, z) d(k \times z) \quad (19)$$

We do not model the entry decisions of entrepreneurs and simply assume that they all start with an initial level of capital k_0 , corresponding to the lowest possible level of capital at which production can take place.²⁴ Equation 19 shows that next period's measure of firms is given by the number of new entrants (summarized by the density $\nu(z'|k = k_0)$, which is degenerate in initial capital) plus the number of incumbents transiting from current state (k, z) to future state (k', z') .

We now have all elements to define a stationary equilibrium for this economy.

Definition 1: A stationary equilibrium is a set of prices $\{w^*, \bar{r}\}$, a set of decision rules, $\{k^{r*}, k^{o*}, l^{d*}, ex^{it*}, N^*\}$, and a distribution of firms $\{\mu^*\}$ such that:

1. Decision rules are optimal
 - l^{d*} , k^{r*} and k^{o*} solve the entrepreneur's problem;
 - N^* solves the representative worker's problem;
2. μ^* is defined recursively by 19 and $\mu^{*'} = \mu^*$;
3. The government budget balances;
4. The wage is market-clearing and aggregate labor demand equals aggregate labor supply: $L^{s*} = L^{d*}$;
5. Demand for rented capital is satisfied at the exogenous interest rate \bar{r} , given the collateral requirement.

²⁴This initial level of capital can be thought of as a fixed cost of entry.

4 Some analytical results

Having specified the production technology, one can easily solve for the optimal hiring decision of an entrepreneur:

$$l = \left[\lambda(l) \left(\frac{\eta}{w} \right) z k^\gamma \right]^{\frac{1}{1-\eta}} \quad (20)$$

where $\lambda(l) = 1 - \tau_p - p_a(l)(\tau_0 - \tau_p)$. This is standard except for λ which is the burden from taxation.²⁵

First best. In order to compare and assess the effect of the two distortions in turn, we examine an economy with no tax liabilities on firms (and thus no audit distortion) and with full-commitment of creditors. This last change results in the elimination of collateral requirements, with a rental market for capital in which entrepreneurs are able to always adjust capital to the desired level, after observing the realization of the productivity shock. The first-best delivers expressions for factor demand that are a function of parameters only:

$$l = z \left[\left(\frac{\gamma}{r} \right)^\gamma \left(\frac{\eta}{w} \right)^{1-\gamma} \right]^{\frac{1}{1-\gamma-\eta}} \quad (21)$$

$$k = z \left[\left(\frac{\gamma}{r} \right)^{1-\eta} \left(\frac{\eta}{w} \right)^\eta \right]^{\frac{1}{1-\gamma-\eta}} \quad (22)$$

The capital-labor ratio in this first-best economy is

$$\frac{k}{l} = \frac{\gamma w}{\eta r} \quad (23)$$

The optimal capital-labor ratio is constant across firms. It is easy to see that a tax on one of the inputs would distort this ratio. It would be harder to obtain an analytical expression of this sort for the benchmark because the investment decision is split over two periods and made, in part, before the realization of the productivity shock. However, the numerical model generates capital-labor ratios that we can compare to this first-best.

In the first-best, the relative demand for labor between any two firms is given by:²⁶

$$\frac{l_i}{l_j} = \left[\frac{z_i}{z_j} \right]^{\frac{1}{1-\gamma-\eta}} \quad (24)$$

It is easy to see that a flat tax rate does not affect this ratio. We also examine the relative demand for labor in the distorted economy:

²⁵Given the functional form for the audit probability (i.e. step-function), the first order condition for labor can be solved in closed hand form.

²⁶A similar expression can be derived for the relative demand for capital.

$$\frac{l_i}{l_j} = \left[\frac{\lambda_i z_i}{\lambda_j z_j} \left(\frac{k_i}{k_j} \right)^\gamma \right]^{\frac{1}{1-\eta}} \quad (25)$$

We derive two important implications from equation 25. First, the institutional distortion acts as a progressive tax. Indeed, the audit schedule decreases the amount of labor employed in any given firm but relatively more in larger firms.²⁷

Second, there is a non-degenerate distribution of resources across firms *within* each productivity ‘class’ (that is, conditioning on a given productivity level) due to each firm’s specific capital accumulation history. It is easy to see from equation 25 that two firms with a different level of capital, $k_i \neq k_j$, but same level of productivity, $z_i = z_j$, will have different numbers of employees.

5 Quantitative Analysis

5.1 Calibration Strategy

Calibration proceeds in two stages. First, some parameter values are assigned using existing estimates from the literature. Then the model is used to estimate the remaining structural parameters. In part this is done by choosing a u-vector of structural parameters, ζ , so that a v-vector of moments simulated from the model, m_s , matches a v-vector of moments from the data, m_d . The distance criterion is a simple quadratic loss function:

$$\hat{\zeta} = \arg \min_{\zeta} (m_s(\zeta) - m_d)' W (m_s(\zeta) - m_d) \quad (26)$$

where W is the variance-covariance matrix from the data. The procedure is based on repeating the following steps until a minimum is reached:

1. Guess a vector ζ of parameters to be estimated;
2. Iterate over the conditional value functions until joint convergence of the optimal functions;
3. Simulate shocks for 10000 entrepreneurs over 100 periods to attenuate the effect of initial conditions;
4. Calculate simulated moments m_s in the exact same way as they are in the data;
5. Compare the simulated moments with those from data m_d ;
6. Update the vector ζ if a minimum has not been reached.

²⁷Result available upon request

Table 5: Parameters

Parameter	Value	Description	Source
Institutional environment: regulation and tax parameters			
l^{ic}	30	Size-threshold into formality	measured from data
τ_0	0.14	Official tax rate	measured from data
τ_p	$\tau_0/2$	Effective tax rate	measured from data
p_a^s	0.53	Probability of audit, firms ≤ 30 employees	measured from data
p_a^l	0.71	Probability of audit, firms > 30 employees	measured from data
Financial development parameters			
ϕ	0.99	Degree of credit market completeness	estimated using model
r	0.21	Rental rate of capital	assigned, external estimate
Technology and production parameters			
δ	0.07	Depreciation Rate	assigned, external estimate
γ	0.28	Share of capital in production	assigned, external estimate
η	0.57	Share of labor in production	assigned, external estimate
ρ	0.99	Persistence	estimated using model
σ_ϵ^2	0.003	Variance White Noise	estimated using model
w	1096\$/yr	Wage	measured from data
φ	0.9	Firms' survival rate	assigned, external estimate
n_i	15	Maximum initial productivity	estimated using model
Preferences			
β	0.97	Intertemporal discount factor	assigned, external estimate
D	1.22	Disutility of labor	estimate using model

Equation (26) is analytically intractable and we cannot rely on standard optimization tools which use gradients to search for a minimum due to the non-convexities involved in the problem. Instead, we use ‘simulated annealing’, a minimization algorithm that proceeds by random search.

5.1.1 Details about parametrization

Table C1 presents the value and origin of all parameters used in the benchmark calibration. Note that we set the length of a time period in our model to be one calendar year.

Institutional environment. The size-threshold into formality, l^{ic} , is based on the regression discontinuity analysis of ‘effective’ tax payments and is set at 30 employees.²⁸ We use the ratio

²⁸This value also corresponds to the maximum number of employees in small firms used in other works. See for example, Steel and Webster (1992), Tybout (2000) and Gauthier and Reinikka (2006).

of tax obligations per sale found in the data to set a flat official tax rate $\tau_0 = 0.14$. Based on our estimates we approximate the effective rate of taxes paid, τ_p , to be half the official rate.²⁹ Audit probabilities for small and larger firms are also taken from the data and set to 0.53 and 0.71 respectively.

Technology. Using US data on 34 industries Basu and Fernald (1997) have estimated total returns to scale for a typical industry, $\gamma + \eta$, to lie between 0.8 and 0.9. We thus make a conservative choice and follow Restuccia and Rogerson (2008) who use $\gamma + \eta = .85$ for the US and attribute parameter values according to capital and labor shares of income, i.e., 1/3 and 2/3, respectively. Note that Soderbom and Teal (2004) find a very similar assignment of income shares for Ghana. We assume a rate of depreciation of $\delta = 0.07$.

Firms' productivity. As in Hopenhayn and Rogerson (1993), there is an identification issue in that the price of output, the productivity shock and the current level of capital enter multiplicatively in the production function. The effect of a high price cannot be distinguished from a high value of the current shock nor from a high level of capital stock. A few assumptions are thus in order. First, we normalize the price of output to one. Second, we follow Hopenhayn and Rogerson (1993) and impose a grid for productivity shocks with $n_z = 20$ log-spaced points as well as a capital grid with 250 log-spaced points where the lower bound is normalized to one.³⁰

We use the relative demand for labor, equation 25, to identify the range of productivity shocks. In order to do so, we use the range of firm sizes as observed in the actual data which is between 1 and 2,000 employees. This gives us the following support for shocks: $z \in [1, 3.8]$.

We assume that productivity shocks for new entrants are initially drawn from a uniform distribution, defined on a subset of the productivity domain of incumbent firms. Therefore the range of the initial distribution, i.e., the subset of points on the productivity grid, is estimated using the model and denoted as n_i . In order to identify this range we add as a target in the loss function (26) the average number of employees in the data: this results in an estimate of $n_i = 15$, meaning that initial productivity assumes values over the lowest 15 points of the productivity grid. For firms that have been in the market for at least one period, productivity shocks follow an AR(1) process:

$$z_t = \rho z_{t-1} + \epsilon_t \tag{27}$$

²⁹Goyette (2009) finds that tax obligations represent 14.4% of sales in 1997. Tax obligations per employee are higher for medium firms which owe an average of US\$4,526 as opposed to US\$1,548 for small firms and US\$3313 for larger firms. Taxes paid represent 7.7% of sales value. Overall, about half of the tax obligations are customarily evaded. 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.

³⁰Our criterion in the sensitivity analysis is that no firms should cluster at the upper bound of the capital grid. This provides us with the smallest possible capital grid which economizes on computational time. Note also that the grid is as in Hopenhayn and Rogerson (1993).

where $0 < \rho < 1$ and ϵ_t is a white noise perturbation. We assume that the distribution of the white noise is a Normal with mean zero and variance σ_ϵ^2 . The AR(1) process is covariance-stationary with mean zero and variance $\sigma_z^2 = \sigma_\epsilon^2 / (1 - \rho^2)$. It can be shown that if ϵ_t is normally distributed then z_t is also normally distributed. we use a method suggested by Kopecky and Suen (2010) to approximate the AR(1) process by a discrete space process. This involves constructing a grid for productivity shocks with n_z points and an upper bound given by:

$$z^{max} = \sqrt{(n_z - 1)}\sigma_z \quad (28)$$

A transition matrix $g(z_{t+1}, z_t)$ is then computed using a parameter $p = \frac{1+\rho}{2}$. We estimate ρ using the simulation-based method described above. In order to identify this parameter the model-to-data discrepancy in average capital-output ratio across firms is added to the loss function (26), as well as the average observed age of a firm.

We pin down the value of σ_ϵ^2 using equation 28. Given that, as described above, the z_{max} is estimated equal to 3.8, we can use the estimated value for ρ to back out σ_ϵ^2 from equation 28.

Exogenous firms' exit. For the survival rate of firms, φ , Roberts and Tybout (1997) estimate that the 1-year exit rate for firms in Chile, Colombia and Morocco are 8.5%, 11.9% and 9.5%, respectively. We assume a middle of the range value of 10% for Uganda and set $\varphi = 0.9$.

Preferences. For discount factors 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). We make a middle of the range choice by setting $\beta = 0.97$. Results are not very sensitive to changes in β .

The wage is normalized to be the same as what we observe in the data (1096 USD per year).³¹ The disutility parameter D is calculated from the first order condition of the worker's problem which gives: $D = 1/N$ where N is the fraction of employed workers in the economy. Using UN data for the employment-to-population for Uganda in 1997, one obtains a value for $D = 1.22$.³²

Financial Development. We set the rental rate of capital $r=0.21$, which is the average commercial bank lending rate for 1997 based on data from the IMF. For the financial development parameter ϕ , other papers have used the external finance-to-GDP ratio to calibrate similar parameters (see Buera and Shin,2008). We do not have a direct counterpart for this measure in our data set and, therefore, we choose to include the share of credit constrained firms (unable to apply for a loan) among the targets in our simulation-based estimation. We expect variation in this dimension conveys at least some information about the extent of credit market access.

³¹There is not much data variation in yearly wages. In the model we assume that the smallest firm in the economy is the marginal employer. Therefore we use equation 20 to derive the marginal product for a firm with 1 employee, the entry level of capital and the lowest shock. This is assumed to be equal to 1096 USD.

³²Employment-to-population ratio for Uganda in 1997 for both sexes is 82%, 88% for men and 76% for women.

Summary. We estimate three parameters using the simulated method of moments. These parameters are: ρ , ϕ and the grid point corresponding to the maximum productivity of new entrants, n_i . Four target moments are included in the loss function: (i) the average number of employees; (ii) the average capital-output ratio; (iii) the share of credit constrained firms in the sample; and (iv) the average age of a firm. We thus have a u-vector of 3 parameters for ζ and a v-vector of four targeted moments.

We use the parameterized model to verify to what extent the input market distortions translate in a missing middle in the size distribution of firms, as well as in capital-labor ratio differences across firms, and growth rates of output, that resemble our empirical observations. We do not target these features explicitly.

5.2 Model validation

5.2.1 Targets

Table 6 contrasts the simulated moments obtained from the calibration with the target moments in the data. The model does a reasonable job in terms of the average size of firms (number of employees). Average age is lower than in the data, but well within the confidence interval in the data.

The average capital-output ratio is also lower in the simulated data than in the actual data, while the share of credit constrained firms is exactly matched. The inability of the model to capture the average capital-output ratio is likely due to the coarse quality of the information we have regarding the degree of credit constraints in the economy: by matching the share of firms that hit their borrowing limit, we implicitly restrict the capital-output ratio. If we had micro data on the debt levels of firms, we could arguably use finer targets, getting a better fit of the average capital-output ratio.

Table 6: Targets

Variable	Benchmark	Data
Average number of employees	135	124 (259)
Average firm's age	10	14 (12)
Average capital/output ratio	1.1	1.3 (3.1)
Share of credit constrained firms	0.53	0.53
Note: Std. dev. in parenthesis		

5.2.2 Non-Targeted Statistics

In this section we examine some model statistics that were not targeted in the calibration. Table 7 summarizes these results.

Table 7: Non-Targeted Statistics

Variable	Firm's size	Benchmark	Data	
		Average	Average	Std. Err.
Share of firms	Small	.51	.48	
	Medium	.08	.20	
	Large	.41	.32	
Average size (employees)	Small	14	14	8
	Medium	60	50	13
	Large	302	334	379
Output growth	All firms	.14	.08	.54
	Small	.06	.02	.43
	Medium	.31	.13	.59
	Large	.20	.17	.67
Share of output	Small	.05	.03	
	Medium	.04	.14	
	Large	.91	.84	
Owned capital per worker	Small	15.6		
	Medium	12.3		
	Large	12.5		
Total capital per worker	Small	55.3	78	129
	Medium	54.9	38	35
	Large	58.2	50	97

Firm Size Distribution. The model generates an equilibrium size distribution of firms with a ‘missing middle’. Examining the share of firms for the three size bins we note that the model tends to overestimate the shares of small and large firms. This is due to two main reasons: (1) we have not modeled the difference in ‘effective’ tax liabilities between medium and large firms; (2) we impose a stark, step-wise jump in audit intensity at the ‘formality’ threshold for small firms, which has no direct counterpart in the data as noted in table 3. However despite these simplifying assumptions, the benchmark produces average sizes that are much in line with those from the data.

Revenue growth. A peculiar feature of the Ugandan data is an unusual pattern in firms’

growth rates, which seems to contrast with other findings in the literature on firms dynamics (Evans, 1987; Hall, 1987). Small Ugandan firms exhibit lower growth rates than their larger counterparts. Interestingly, the quantitative model delivers patterns in growth rates that qualitatively mimic the pattern observed in the data for small and medium firms (see table 7). The magnitude of the growth rates for small and medium firms are quantitatively larger than what we estimate for Uganda. In table 8 below we report reduced-form regressions of revenue growth on the number of employees. Only firms with less than 100 employees are used for both the actual and the simulated data as we want to focus on small and medium firms.

Table 8: Revenue growth and Number of Employees: OLS Estimates

Dependent Variable: Growth rate of Revenues (1996-1997) ^a		
	Data	Benchmark
# of Employees	.005 (.002)	.002 (.001)
Constant	-.06 (.06)	.03 (.03)
R^2	.06	.05
# of observations	157	169

Note: Std. dev. in parenthesis

Capital-labor ratios. We also examine the patterns in capital-labor ratios, using both total capital and owned capital. Notice that one can interpret the share of rented capital, over total capital, as a measure of outstanding firm’s debt. Unfortunately, given our data limitations, we have no empirical counterpart for this measure.

Table 7 shows that the average capital-labor ratio, using owned capital, is substantially greater for small firms than for medium firms: this is a reflection of both the collateral constraints and the input choice distortion due to the audit probability change. Notice that the ratio for large firms goes back up and is larger than for medium firms. However, using total capital, the difference between small and medium firms is greatly reduced, and is smaller than what we observe in the data. This discrepancy is likely the result of the limited information we have about debt-to-equity ratios of firms and the fact that we use a coarse target for modeling credit constraints.

Some interesting results can be shown through a graphical illustration. Figure 4 and 5 show, respectively, owned and total capital per worker in the benchmark economy. In our model there is no input price dispersion across firms, therefore each dot represents a cluster of firms with identical productivity, consistently with the discussion in section (4): this simplification however makes it much easier to infer the basic mechanisms at work. First, at the bottom of the graph one can notice those firms which, after experiencing good productivity shocks, are hitting the

borrowing limit, as they line up in order of decreasing capital intensity. The constraint on external finance is especially strong for small establishment that start with little capital. On the other hand, small firms with higher levels of collateral clearly exhibit capital intensities which are at least as large as medium firms, with the ratio become larger as they approach the formality threshold (30 employees) and start substituting capital for labor. The interaction of constrained external finance and uneven taxation results in a large dispersion of capital intensities among small firms, as observed in the data, with a clustering of firms right at the size-threshold of 30 employees, where the audit intensity increases. This clustering of firms at the threshold is even more obvious in figure 6 below, where we plot the capital intensities in the case of unconstrained access to external finance: in that particular experiment there is no dispersion within each productivity group, so it is possible to see in a very direct way how small firms close to the 30 employees threshold ramp up their capital-labor ratios in order to avoid the switch to the formal, higher tax status. Finally, there is a visible gap in the size distribution of firms between 30 and 50 employees.

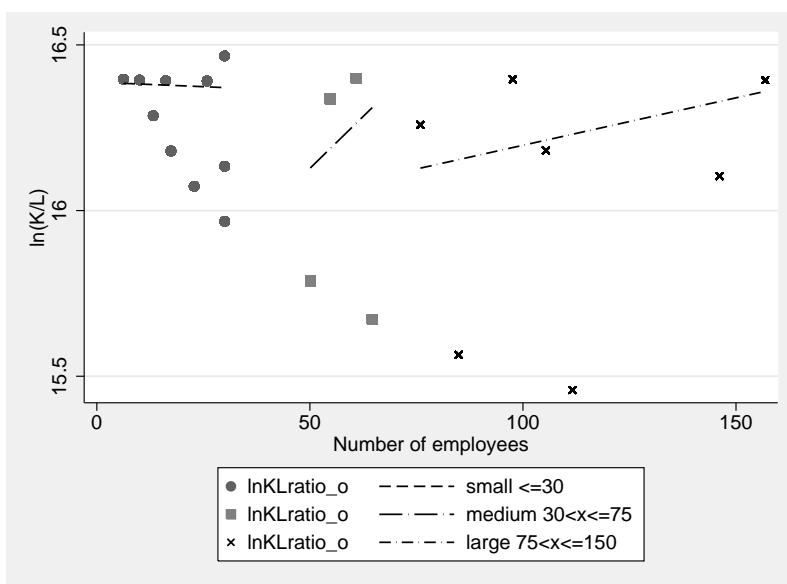


Figure 4: Simulated Capital-Labor Ratios v.s. Firm Size (Owned Capital)

5.2.3 Counterfactual Analysis

In this section we examine the qualitative and quantitative implications of the observed distortions by comparing the benchmark equilibrium to 4 alternative scenarios. We start by examining the first-best allocation. In this scenario, corporation taxes are completely removed and access to credit is unconstrained. This exercise is performed under two different assumptions about

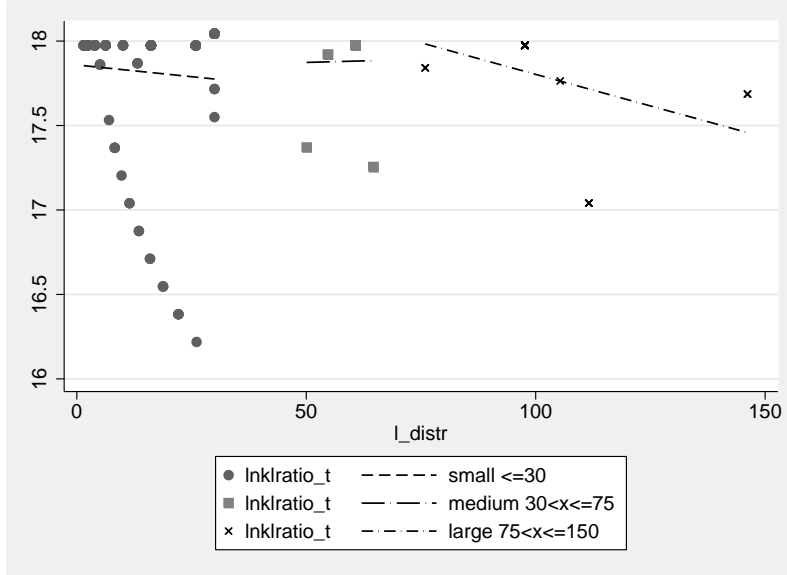


Figure 5: Simulated Capital-Labor Ratios v.s. Firm Size (Total Capital)

the price of credit (one based on average observed commercial rates, the other assuming a rate similar to that observed in more developed economies).

In order to assess the relative importance of different distortions in explaining the efficiency deviation from the first best, we do the following: (i) remove taxes while keeping the level of financial development ϕ at the benchmark level; (ii) keep the existing corporate tax and audit scheme, but set the financial parameter ϕ to a value such that entrepreneurs can always rent the desired amount of capital (this is true when $\phi = \phi^{max} = r/(1 - \delta) + 1$). This latter exercise corresponds to the benchmark economy with unconstrained access to credit.

Finally, we analyze what happens when the credit environment is the same as in the benchmark economy, but we flatten the audit intensity so that all firms face the same audit intensity. In this experiment we are careful to set the identical audit probability to a level such that tax revenues are the same as in the benchmark.

The experiments are first run under the assumption that equilibrium in the labor market is achieved only through wage adjustment, with aggregate labor supply held constant. This first set of experiments focus on the effects of pure labor reallocation. It is likely, however, that a labor supply response would be elicited by changes in the economic environment and in firms' behavior. Therefore we also report results for experiments in which both wage and labor supply are allowed to adjust in equilibrium: in particular, we assume that the wage increase in equilibrium corresponds to one half of the wage increase observed in the first set of experiments (the ones with constant aggregate labor supply), and that the remaining labor market clearing

is borne out by changes in aggregate labor supply.³³

Table 9 and 10 summarize the percentage change from the benchmark for key variables. Of course, the aggregate responses are much larger when we allow for labor supply responses. In the discussion that follows we focus on output per worker and assess the efficiency losses associated to different distortions. Given the assumption of DRS technology, per worker efficiency is always higher in the experiments in which aggregate labor supply is held constant. Notice that the labor adjustment in the first-best experiments is quite large, and correspond to a large labor elasticity. Therefore the results regarding changes in capital and output per worker can be considered as a lower bound of the true effect of the policy.³⁴

³³We have also performed, but not reported, experiments under the extreme scenario in which wages remain constant at the benchmark level and that all the labor market adjustment is borne out by changes in aggregate labor supply.

³⁴We have chosen to present these results so to provide an illustration of an extreme response in labor supply. Given our data, we have no basis to guess what the true aggregate labor elasticity would be in a developing country like Uganda.

Table 9: Experiments with inelastic aggregate labor supply

Experiment	1a.First-Best r=0.21	1b. First-Best r=0.05	2. No corporate tax r=0.21	3a. Unrestricted access to credit r=0.21	3b. Unrestricted access to credit r=0.05	4. Flat audit intensity r=0.21
Change in:			(% change from the benchmark)			
wage	26	122	20	4.9	85	0
aggregate output	10	94	5.1	4.9	85	0
aggregate capital	31	869	20	9	707	0
Capital per employee			Value in equilibrium			
small firms	77.6	574.5	64.4	64.9	480.2	54.2
medium firms	77.6	574.5	67.0	64.7	478.5	56.5
large firms	77.6	574.5	69.5	64.7	478.5	58.3
firms of different sizes			(% change from the benchmark)			
small firms	-27	-27	-13	-13	-13	-13
medium firms	88	88	89	-2	-2	89
large firms	16	16	-1.5	16	16	-1.5
credit constrained firms	-100	-100	15	-100	-100	5

Table 10: Experiments with aggregate labor adjustment

Experiment	1a. First-Best r=0.21	1b. First-Best r=0.05	2. No corporate tax r=0.21	3a. Unrestricted access to credit r=0.21	3b. Unrestricted access to credit r=0.05
Change in:	(% change from the benchmark)				
wage	13	61	10	2.5	43
aggregate labor	68	364	48	12	245
aggregate output	66	553	43	14	392
aggregate capital	97	3,154	63	19	2,049
Capital per employee	Value in equilibrium				
small firms	69.7	417.0	57.4	63.6	375.3
medium firms	69.7	417.0	62.0	63.2	369.1
large firms	69.7	417.0	62.9	63.2	369.1
firms of different sizes	(% change from the benchmark)				
small firms	-40	-62	-24	-13	-40
medium firms	74	45	73	-2	-16
large firms	35	69	16	16	53
credit constrained firms	-100	-100	42	-100	-100

Experiments 1a and 1b: First-Best In the first-best, corporate taxes are removed and firms can always rent the desired amount of capital after observing their productivity shock. In experiment 1b, in tables 9 and 10, we use a rate of rental of capital of the magnitude observed in developed economies, say 5%. This brings about the highest changes in aggregate variables. When only the wage is adjusted (table 9), this results in a 2-fold increase in output per worker and a more than doubling of the wage rate. Note that if we let both labor and wages adjust, the increase in output per worker is of the order of 52%, with a 61% increase in wages. The value of capital-labor ratios in the first best experiments are identical across firms, exactly as the theory predicts in the absence of price dispersion (see equation 23). The change in the size distribution of firms are such that there are more medium and large firms and, by construction, no firms face credit constraints in these experiments.

In experiment 1a the interest rate is held at the benchmark level: when we allow for aggregate labor adjustment (see table 10), aggregate capital per employee increases by 43% however output per employee is virtually unchanged. In contrast, when we only let the wage adjust, aggregate capital and output per worker increase by 31% and 10%, respectively (see table 9). Wage changes are also smaller, ranging between 13 and 26%.

Overall these experiments suggest that large output increases are likely to be associated to changes in the cost of external finance, given unrestricted access to credit and no corporate taxes. Large differences in productivity per worker can be explained by the differences in the price of credit: the result would be even stronger if we had assumed a price setting mechanism which allows for price discrimination in equilibrium, as it probably happens in reality. Unfortunately we have no data to model such price dispersion and we have chosen to focus on the simple degenerate case.

Experiment 2: No corporate taxes In this experiments credit conditions are the same as in the benchmark, but we remove corporate taxes. From table 9 we note that, keeping aggregate labor constant, the wage goes up by 20%. The increase in aggregate capital and output are 20% and 5.1% respectively. Importantly, the capital-labor ratio for small firms is now lower than that of medium firms. This should come as no surprise given our previous discussion of both data and model, and it confirms that changes in effective taxation are responsible for the ramping up of capital intensity just before the 30 employees' threshold. Also, as seen from the percentage changes in the share of small and medium firms, several firms which were small in the benchmark appear to switch to the medium size bin. These firms are those which experienced positive productivity shocks and would scale up their total share of external finance if they could, so we also observe a 15% increase in the share of credit constrained firms. When we allow labor to adjust (table 10) we observe an increase in aggregate labor (48%) that offsets the increase in

output (43%), due to the decreasing aggregate returns to scale. Overall this experiments confirm that tax distortions, rather than credit constraints, can explain the missing middle of the size distribution and the relatively high capital intensity of small firms around the 30 employees' threshold. However the experiments also suggest that the efficiency losses associated to such distortions only explain a small fraction of the total efficiency deviation from the first-best.'

Experiment 3: Perfect Capital Markets These experiments maintain the tax and audit setup of the benchmark economy, and examine the equilibrium outcomes with unconstrained access to credit. Under the assumption that $r = 0.21$, table 9 shows that the wage must increase by 5% to keep aggregate labor as in the benchmark. Aggregate capital and output increase by 9% and 4.9%, respectively. Note that the average capital-labor ratio is still marginally higher for small firms than for medium firms, confirming the importance of the regulation distortion for this particular feature of the data. However, consistent with the theory, we observe constant capital-labor ratios across medium and large firms. Finally, the main change in the size distribution is an increase in the share of large firms, because the wage does not increase as much as in experiment 2. Since these firms undertake most of the production, releasing credit constraints will entail a larger increase in aggregate output.

Lowering the interest rate (experiment 3b. in table 9), brings about much larger changes. Output per worker goes up by 85%. In fact, the combination of unrestricted access and lower cost of external finance accounts for most of the efficiency discrepancy between benchmark and first-best. This happens because the change in interest rate is especially beneficial to firms with a high shadow value of capital, which are the ones experiencing high productivity spells. The effects when we let labor adjust are similar but of smaller magnitude (table 9) and can be thought of a lower bound of the effects considered.

Finally, in figure 6 we clearly see that around the level of 30 employees there is a cluster of firms which substitute capital for labor. As predicted by the theory, we note that there is no more dispersion within a productivity class when moving from the benchmark to an economy with perfect capital markets. Dots in figure 6 represent clusters of firms making the same optimal choice of labor and capital, conditional on their productivity. Of course, in reality input price dispersion would result in larger capital/labor dispersion. However we see this experiment as a compelling illustration of the mechanism driving up capital labor ratio for small 'informal constrained' firms!

Experiment 4: Flat Audit Intensity For the case of inelastic labor supply, we also perform an experiment in which we flatten the audit probability, adjusting it so to keep aggregate

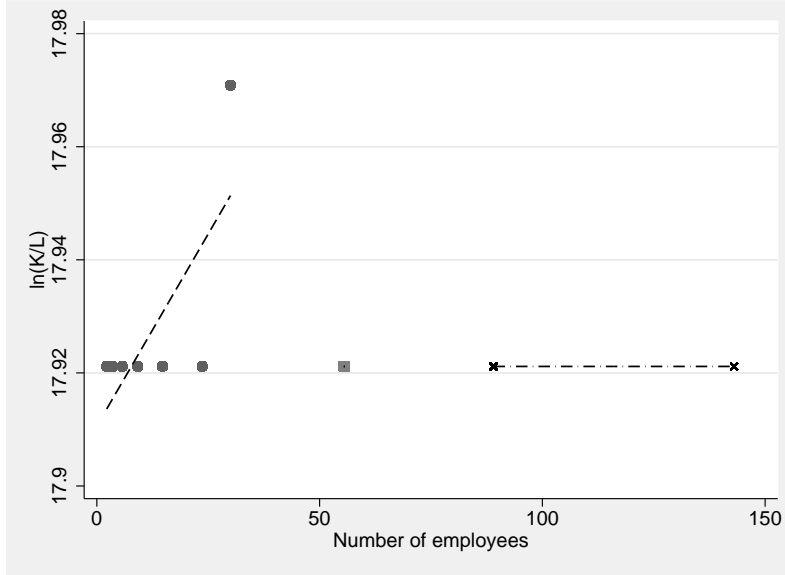


Figure 6: Simulated Capital-Labor Ratios v.s. Firm Size for Experiment 3

tax revenues at their benchmark level.³⁵ The purpose of this exercise is to assess the extent to which this change would result in labor reallocation and changes in growth patterns. As it can be seen in table 9, an even effective tax yields no change on aggregate wage, capital and output. We do observe a reallocation of firms from the small size bin to the medium size bin of the same order as in experiment 2 when the wage is adjusted. Capital-labor ratios for small firms are now *lower* than for medium firms. The ‘missing middle’ in the firms’ size distribution disappears, suggesting that it is purely a by-product of the distortions in the tax system and that it is associated to negligibly small losses in aggregate efficiency, unless taxes are eliminated as in experiment 2. We conclude that the ‘missing middle’ simply reflects the optimal responses of forward looking firms to the tax and auditing environment, and that such responses are quite effective in reallocating resources so to minimize the impact of the distortion.

Discussion Overall we interpret these results as suggesting that credit market distortions are likely to play a major role when explaining low levels of productivity per worker in developing countries like Uganda. For the case of pure labor reallocation, 9 shows that removing the tax distortion (experiment 2) would reduce the efficiency gap between benchmark and the first-best with low $r=0.05$ (experiment 1b) by 5%. The introduction of perfect capital markets (experiment 3a) would reduce the gap by a further 5%. But it really is the cost of capital (experiment 3b) that goes the longest way in closing the gap between the benchmark and the low interest first-best, by covering the remaining 85%.

³⁵The overall audit intensity that yields the same tax revenues as in the benchmark is $p_a = 0.7$.

Also, all the experiments and in particular the results from experiment 4, suggest that the ‘missing middle’ in the size distribution of firms is due to the regulation distortion, rather than imperfect credit markets. The ‘missing middle’ in the firms’ size distribution is a common observation in developing countries. Our analysis suggests it might be the by-product of institutional distortions in the tax system (such as labor market regulation, uneven auditing or large differences in the taxation of capital and labor) but we find only relatively small losses in aggregate efficiency associated to the missing middle: in fact, the missing middle seems to be the proof that input choices can be adjusted to effectively reallocate resources and minimize the impact of the distortions.

6 Conclusion

We analyze the resource allocation and efficiency implications of large input market distortions due to uneven taxation and imperfect credit markets. The distortions are measured using microdata information from a sample of Ugandan firms. We calibrate and numerically simulate an amended model of firms’ growth in to measure the equilibrium effects of the distortions considered. The model does a reasonably good job at replicating some unusual, non-targeted, data patterns. Through counterfactual experiments we find that efficiency losses in output per worker in the benchmark economy, vis-a-vis an undistorted first best are very large, ranging between a lower bound estimate of 52% and an upper bound estimate of 94%. Perhaps more surprisingly, we find that most of the efficiency loss (90% or more) is attributable the restricted access to and high cost of external credit.

We also find that the ‘missing middle’ in the firms’ size distribution is due to the optimal input responses to tax distortions, rather than credit constraints, but that such responses are associated to relatively small losses in aggregate efficiency. In fact the ‘missing middle’ simply reflects the input adjustment made by forward looking firms given the existing tax and auditing environment. We find that such responses are quite effective in reallocating resources so to minimize the impact of the distortion.

References

- Acemoglu, D., Johnson, S., Robinson, J., 2001. The colonial origins of comparative development: An empirical investigation. *The American Economic Review* 91 (5), 1369–1401.
- Alfaro, L., Charlton, A., Kanczuk, F., 2007. Firm-size distribution and cross-country income differences. manuscript, Harvard Business School.

- Arellano, C., Bai, Y., Zhang, J., 2008. Capital Structure and Contract Enforcement. Working Paper.
- Banerjee, A., Duflo, E., 2005. Growth theory through the lens of development economics. *Economics Handbook of Economic Growth 1, Part A*, 473–552.
- Banerjee, A., Duflo, E., 2008. Do firms want to borrow more? Testing credit constraints using a directed lending program. MIT Working Paper Series.
- Basu, S., Fernald, J., 1997. Returns to scale in US production: Estimates and implications. *Journal of Political Economy* 105 (2), 249–283.
- Buera, F., Kaboski, J., Shin, Y., 2009. Finance and development: A tale of two sectors. NBER Working Paper.
- Buera, F., Shin, Y., 2008. Financial Frictions and the Persistence of History: A Quantitative Exploration. Manuscript, Washington University in St. Louis.
- de Soto, H., 1989. *The Other Path: The Invisible Revolution in the Third World*. London. IB Tauris.
- Easterly, W., Levine, R., 2001. What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models. *The World Bank Economic Review* 15 (2), 177.
- Evans, D., 1987. The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. *The Journal of Industrial Economics* 35 (4), 567–581.
- Garcia-Santana, M., Pijoan-Mas, J., 2010. Small Scale Reservation Laws and the Missallocation of Talent. Working Paper.
- Gauthier, B., Gersovitz, M., 1997. Revenue erosion through exemption and evasion in Cameroon, 1993. *Journal of Public Economics* 64 (3), 407–424.
- Gauthier, B., Reinikka, R., 2006. Shifting tax burdens through exemptions and evasion: an empirical investigation of Uganda. *Journal of African Economies* 15 (3), 373.
- Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. *Journal of Applied Econometrics* 8, 85–118.
- Goyette, J., 2009. Determinants of the Size Distribution of Firms in Uganda. Working Paper, UBC.

- Guner, N., Ventura, G., Xu, Y., 2008. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics* 11 (4), 721–744.
- Hall, B., 1987. The relationship between firm size and firm growth in the US manufacturing sector. *The Journal of Industrial Economics* 35 (4), 583–606.
- Hall, R., Jones, C., 1999. Why Do Some Countries Produce So Much More Output Per Worker Than Others?*. *Quarterly Journal of Economics* 114 (1), 83–116.
- Hopenhayn, H., 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60 (5), 1127–1150.
- Hopenhayn, H., Rogerson, R., 1993. Job turnover and policy evaluation: A general equilibrium analysis. *The Journal of Political Economy* 101 (5), 915–938.
- Hsieh, C., Klenow, P., 2007. Misallocation and Manufacturing Productivity in China and India. NBER Working Paper 13290.
- Klenow, P., Rodriguez-Clare, A., 1997. The neoclassical revival in growth economics: Has it gone too far? NBER macroeconomics annual 12, 73–103.
- Kopecky, K., Suen, R., 2010. Finite state Markov-chain approximations to highly persistent processes. *Review of Economic Dynamics* 13 (3), 701–714.
- Liedholm, C., Mead, D., 1987. Small scale industries in developing countries: empirical evidence and policy implications. *International Development Papers*.
- Little, I., Mazumdar, D., Page, J., 1987. Small manufacturing enterprises: A comparative analysis of India and other economies. A World Bank Research Publication.
- McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142 (2), 698–714.
- Midrigan, V., Xu, D., 2009. Finance and Misallocation: Evidence from Plant-Level Data. Working Paper.
- North, D., et al., 1981. *Structure and change in economic history*. Norton New York.
- Prescott, E., 1998. Needed: A theory of total factor productivity. *International Economic Review* 39 (3), 525–552.
- Restuccia, D., Rogerson, R., 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11 (4), 707–720.

- Roberts, M., Tybout, J., 1997. Producer turnover and productivity growth in developing countries. *The World Bank Research Observer* 12 (1), 1.
- Sleuwaegen, L., Goedhuys, M., 2002. Growth of firms in developing countries, evidence from Cote d'Ivoire. *Journal of Development Economics* 68 (1), 117–135.
- Smith Jr, A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, 63–84.
- Soderbom, M., Teal, F., 2004. Size and efficiency in African manufacturing firms: evidence from firm-level panel data. *Journal of Development Economics* 73 (1), 369–394.
- Steel, W., 1993. Small Enterprises in Indonesia: Role, Growth, and Strategic Issues, DSP Working.
- Steel, W., Webster, L., 1992. How small enterprises in Ghana have responded to adjustment. *The World Bank Economic Review* 6 (3), 423.
- Tybout, J., 2000. Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic literature* 38 (1).