Distortions, Efficiency and the Size Distribution of Firms

Jonathan Goyette* Giovanni Gallipoli

Université de Sherbrooke University of British Columbia

April 14, 2015

Abstract

We develop a model of firms' growth in which the tax and credit environments act as selection mechanisms. Such a model, parametrized and validated using a variety of data restrictions, can rationalize observations about input choices and size patterns typical of many developing countries. Using counterfactual experiments, we show that firms' optimal responses to the tax environment are effective in reducing efficiency losses. As a consequence, tax distortions only account for 13% of the gap in output per worker between an undistorted economy and the benchmark. Credit constraints account for 44% of this gap. However, the interaction between the cost of capital and credit constraints appears to be the most important source of misallocation and can explain up to 85% of the difference in output per worker between the benchmark and first-best.

JEL codes: C15, E27, H26, L11, O41, O47

Keywords: Growth, Size Distribution of Firms, Financial Frictions, Taxation, Calibration

*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

Title: Distortions, Efficiency and the Size Distribution of Firms

Abstract

We develop a model of firms' growth in which the tax and credit environments act as selection mechanisms. Such a model, parametrized and validated using a variety of data restrictions, can rationalize observations about input choices and size patterns typical of many developing countries. Using counterfactual experiments, we show that firms' optimal responses to the tax environment are effective in reducing efficiency losses. As a consequence, tax distortions only account for 13% of the gap in output per worker between an undistorted economy and the benchmark. Credit constraints account for 44% of this gap. However, the interaction between the cost of capital and credit constraints appears to be the most important source of misallocation and can explain up to 85% of the difference in output per worker between the benchmark and first-best.

JEL codes: C15, E27, H26, L11, O41, O47

Keywords: Growth, Size Distribution of Firms, Financial Frictions, Taxation, Calibration

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 exploits information on two measurable input distortions in Uganda to back out the optimal 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 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

[1] Along these lines, a new applied literature is flourishing. Hsieh and Klenow (2009) 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 (2014) examine the costs of size-dependent policies distorting production scale. Buera and Shin (2013), Midrigan and Xu (2014) and Buera et al. (2011) examine the effect of financing constraints on TFP.

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 follows from Author (yearb) and Author (Unpublished resultsa) and introduces size-dependent tax liabilities. Empirical evidence suggests that tax liabilities jump in a discrete fashion when the number of employees grows above a certain threshold and that the uneven tax auditing results in firms with more employees paying higher taxes. The interaction of limited credit market access and size-dependent tax liabilities leads to endogenous dispersion in capital-labor ratios, revenues and profits for firms with similar idiosyncratic productivity.

The model is parametrized in two steps: first some parameters are assigned using external estimates. In particular, the parameters for the fiscal environment and technology are taken from the existing literature. Then, we estimate some parameters using the model.[2] The parametrized model is benchmarked using a variety of data restrictions and then used to perform some counterfactual exercises.

The model can, to a large extent, reproduce peculiar features of the Ugandan data that were not explicitly targeted. In particular, the model reproduces patterns in growth rates and capital-labor ratios across firm size which are comparable to the actual data. The Ugandan data present an unusual pattern in firms' growth rates as small firms exhibit lower growth in revenues than their larger counterparts. Moreover, Ugandan firms' capital-labor ratios also exhibit a very peculiar pattern in that small firms have more capital per employee than larger firms. Interestingly, the model is able to replicate these two facts without explicitly targeting for them. In addition, the model generates a clustering of small firms right before the sharp increase in audit intensity. 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'. These choices generate a gap in the size distribution of firms (as measured by the number of

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

employees) and result in the so-called 'missing middle' phenomenon, typical of many developing countries.[3]

We examine the theoretical implications of the model by sequentially removing the different distortions. We find that removing the tax distortion would close the gap in output per worker between the benchmark and an undistorted economy by 13%. Removing capital rental constraints generates larger gains and explains roughly 44% of the gap. However, the interaction between the rental rate of capital and access to credit seems to explain the bulk of the gap by accounting for 85% of the output difference between the benchmark and the undistorted economy.

We find that both distortions exacerbate the thickness of the left tail of the size distribution of firms. On the one hand, the 'missing middle' in the size distribution of firms seems to be a by-product of the audit distortion in the tax system and is associated to smaller losses in aggregate efficiency than the financial distortion. Our results suggest that the 'missing middle' reflects the optimal responses of forward looking firms to the tax environment, and that such responses are quite effective in reallocating resources so to minimize the impact of the distortion. On the other hand, credit constraints play an important role in shaping the overall size distribution of firms. Indeed, higher levels of credit constraints are associated with a thicker left tail as well as fewer, but larger, firms in the right tail of the size distribution. In other words, credit constraints are associated with size distributions of firms that are more skewed to the right.

Our results are related to a growing literature on the effects of credit and institutional distortions on efficiency and the size distribution of firms in developing countries (see discussion below). Countries do differ in the type of distortions, and in the reasons underlying the shape of their size distribution of firms and aggregate efficiency. However, many developing countries

[3]Many papers in the development literature argue that developing economies exhibit too few medium firms relative to developed economies like the US (Tybout, 2000; Sleuwaegen and Goedhuys, 2002; Author, yearb). More recently Hsieh and Olken (2014) have argued, based on observations in Indonesia, India and Mexico, that bimodal distributions do not constitute the norm in developing countries and that on top of missing medium firms, large firms seem also to be missing. In other words, size distribution of firms in developing countries exhibit thicker left tails compared to developed economies, a fact corroborated by Alfaro et al. (2007).

share some crucial traits: their tax systems are works in progress and share many features with the one in Uganda. For example, over the last few decades, many countries have adopted a value-added tax structure with a tax registration threshold as in Uganda (Keen and Lockwood, 2010). Moreover, credit markets are often in the process of developing and low income countries often struggle with strong constraints in both quantity and price of credit (Cihak et al., 2013). Understanding the way these distortions interact, and the mechanism linking them to the size distribution of firms and to aggregate efficiency, allows one to draw more general inference and provide much needed insights on the efficacy of policies in developing countries.

1.1 Related Literature

Fiscal Environment It is often argued that tax administrations in developing countries target larger firms more intensively because tax collection is expected to be greater than enforcement costs (de Soto,1989; Gauthier and Reinikka,2006). Based on firm-level data from Uganda, Author (yearb) shows that the auditing strategy of tax officials changes around a certain threshold corresponding to about 30 employees. Based on theses empirical facts, Author (Unpublished resultsa) examines, using a model à la Hopenhayn (1992), whether a change from the effective tax threshold, which is determined by the number of employees, to the official tax threshold which is based on total sales, impacts firm level productivity as well as aggregate productivity in Uganda and shows that there could be as much as a 16% gain in aggregate productivity in using the official tax threshold. This paper builds on these results and examines the relative importance of the tax distortion compare to credit constraints.

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 (2013) 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. (2012) 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. (2011) 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 (2014) 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. Our results on the effect of credit constraints lie in between the results of these two subsets of papers.

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

[4] However, 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.

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. For a detailed discussion about the data and summary statistics see Author (yearb).

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 the literature; for example, Tybout (2000), Gauthier and Reinikka (2006), Soderbom and Teal (2004) and many other papers in the development literature use a similar classification of firm size.[6]

For the sake of completeness, we list some noteworthy features of the Ugandan data used to build the theoretical model in the next section: (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. These facts are documented in Author (yearb) and we refer the interested reader to the original article. There is however another fact that we use to build the model in this paper: small firms finance most of their investment with their own cash flow, while larger firms have access to credit.

Cash flow constraint In the 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.

[5] Namely, Kampala, Jinja–Iganga, Mbale–Tororo, Mukono and Mbarara.

[6] Results 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.

Unfortunately the data does not provide information about the outstanding debt level of firms, or the ratio of equity to debt but it gives information about the main sources of financing. We find that access to credit is limited for small firms, using correlations between different sources of financing and firm size as shown in Table 1.[7] We find in Table 1 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.

[Table 1 about here]

There are three peculiar features of the data which might be the direct consequences of the two aforementioned distortions: (1) the Ugandan firms' size distribution has relatively thick left tail 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. For feature 1, we refer the interested reader to Author (yearb). We present some results for the two other features in what follows.

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

$$FirmsGrowth = ln(Sales1997) - ln(Sales1996) \tag{1}$$

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 2: 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.

[Table 2 about here]

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)

[7]Sources 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.

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.[8]

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 1 reports the correlation between capital-labor ratios and firm size for all firms between 1 and 150 employees.[9] Using a t-test, we can show that the average capital-labor ratio for the first size group (blue dots) is larger than for medium firms (green dots).[10] 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.

3 A Model of Firm's Growth

3.1 The benchmark economy

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

3.1.1 Entrepreneurs

In equilibrium there is a measure 1 of risk-neutral entrepreneurs who produce a homogeneous consumption good. Entrepreneurs make production decisions – that is, capital rental and labor hiring – based on: (1) their current idiosyncratic productivity shock, z_t which is drawn from

[8]The 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.

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

[10]t-test results available upon request.

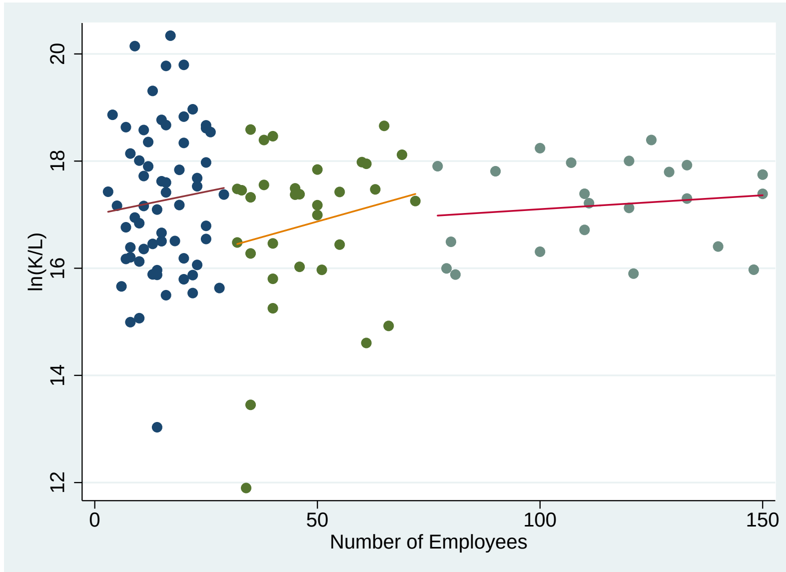

Figure 1: Capital-Labor Ratios v.s. Firm Size

an initial distribution $\nu(k, z)$ and evolves according to a transition function, $g(z_{t+1}, z_t)$; (2) their specific stock of owned capital carried over from the previous period $k^o \in (0, k^{up}]$; (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 developing economies, 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 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.[11]

Ugandan data suggest that, on average, 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

[11]For example,Rosenzweig and Wolpin (1993) argue that, in India, bullocks serve as production capital but can also be sold to smooth consumption due to income shocks. Moreover, using data on 14932 firms in 28 countries in Africa and Latin America, Author (Unpublished resultsb) shows that a large share of the sample (64%) finances working capital and fixed assets with internal funds or retained earnings. Moreover, these firms do not apply for loans due to reasons linked to financial underdevelopment and rely on retained earnings and profits to finance their activities.

tax dues. However, in the event of being audited, they are forced to repay the whole surplus from tax evasion. 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 exits and is replaced by a mass of new entrants in the next period.

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:

$$F(z,k,l) = zk^{\gamma}l^{\eta} \tag{2}$$

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 total tax liabilities $T(l_t)$ which varies with firm's size:

$$T(l_t) = p_a(l_t)T_t^e + T_t^p \tag{3}$$

where $p_a(l_t)$ is the probability of being audited, which varies with firm size such that and $p_a^s < p_a^l$ where s stands for small firms and l for larger firms.

An audited entrepreneur has to repay the amount of taxes evaded T^e.[12] T^p is the amount of taxes actually paid by an entrepreneur. Note that $T^e + T^p = T^0$ where T^0 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 pro-

[12]This 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. See Author (Unpublished resultsa) for further details on and a more refined modeling of the tax distortion.

ductive 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.[13] 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. (2011).

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 were 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:[14]

$$F(z, k, l) - wl + (1 - \delta)k^o - rk^r \tag{4}$$

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

[14]For simplicity we abstract from taxation and fixed costs, which would cancel out anyway, while solving for the condition on rented capital.

where δ is the depreciation rate of capital. Hence, equation 4 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:

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

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:

$$k^r \leq \frac{\phi(1-\delta)}{1+r-\delta-\phi(1-\delta)} k^o \tag{6}$$

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^o_{t+1}. The objective function of a risk neutral entrepreneur is:

$$\max_{c_t, k^o_{t+1}} E_0 \sum_{t=0}^{\infty} (\varphi\beta)^t c_t \tag{7}$$

subject to

$$c_t + k^o_{t+1} = \pi_t \tag{8}$$

$$c_t \geq 0$$

$$k^o_{t+1} \geq 0$$

$$k_0 \;\; \text{given}$$

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

$$\pi_t = F(z_t, k_t, l_t) - w_t l_t + (1-\delta)(k^o_t + k^r_t) - (1 + r_t - \delta)k^r_t - T(l_t) - c_f \tag{9}$$

where c_f is a fixed cost of production and k^r is subject to the constraint defined in equation 6. Given current shock z_t and current level of owned productive capital, k^o_t, an entrepreneur chooses how much to produce by choosing whether to rent additional capital k^r_t 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.[15]

[15] Given that the difference in audit intensity between medium and large firms is not large, we do not introduce

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:

$$\sum_{t=1}^{\infty} \beta^t \left[u(c_t) - d(n_t)\right] \tag{10}$$

subject to

$$c_t \leq w_t n_t \tag{11}$$

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 diversify idiosyncratic risks through employment lotteries which take the form of extended family ties or informal credit in developing countries. Workers are aggregated through a representative agent with preferences given by:

$$\sum_{t=1}^{\infty} \beta^t \left[u(c_t) - DN_t\right] \tag{12}$$

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.

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

$$G \leq T_t^{gvt} = \int \left[T_t^p + p_a(l_t) T_t^e\right] \mu^*(k^o, z) d(k \times z) \tag{13}$$

an exogenous size-threshold at the upper bound of the medium size bin. We choose to keep the model simple and assume an identical audit intensity for medium and large firms.

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

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 6, 8, 9. Given the current number of employees and audit probability, one can define the conditional value functions associated with each formality status $s = \{iu, ic, f\}$ as:

$$V^s(k, z) = \max_{k'} \left\{ c^s + \varphi\beta \int_{z'} \max \left[V^{iu}, V^{ic}, V^f \right] dg(z', z) \right\} \tag{14}$$

where the integral represents the continuation value, defined as maximum expected utility over future formality choices.

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 is less or equal to zero. The measure of firms with current state (k, z) who choose to exit the market is denoted as $ex(k, z)$. We can now define the 'unconditional' value function as the upper envelope of the conditional value functions 14:

$$V(k, z) = \max_{ex,k'} V^s(k', z') \tag{15}$$

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*}, $k^{o\prime*}$ and $exit^*$.

Finally, for an entrant, expected discounted profits are given by:

$$V^e(k,z) = \int V(k,z)\nu(z'|k=k_0)dz' \tag{16}$$

Free-entry implies that:

$$V^e = c_e \tag{17}$$

where c_e is an entry cost. We assume that there exists a frictionless agency that gathers the capital of the exiting firms and re-sells this capital to entering firms and that the cost of the initial capital, k_0, which is the same for all entrants, is included in the entry cost.

3.2.2 Definition of equilibrium

Probability measure $\mu(k,z)$ describes the state of the industry in a given period of time.[16] 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 with probability $(1-\varphi)$. The total measure of exiting firms, ξ, is defined as:

$$\xi = \int \left[\varphi ex(k,z) + (1-\varphi)\mu(k,z)\right] d(k\times z) \tag{18}$$

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:

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

where $G(k,z;k',z')$ represents a transition function mapping current productivity states into future productivity states. 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

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

(k', z').

Let us now define a stationary equilibrium for this economy.

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

1. Decision rules are optimal

   - *l^{d*}, k^{r*} and k'^{*} 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 r, given the collateral requirement.

4 Some analytical results

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

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

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.

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:

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

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

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

$$\frac{k}{l} = \frac{\gamma w}{\eta r} \tag{23}$$

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:

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

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:

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

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

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.

The third implication follows from the second one. If credit constraints are more important for smaller firms due to collateral requirements, this is reflected on the thickness of the left tail of the size distribution of firms.

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:

$$\widehat{\zeta} = \arg\min_{\zeta} \left(m_s(\zeta) - m_d\right)' W \left(m_s(\zeta) - m_d\right) \tag{26}$$

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.

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.[17]

5.1.1 Details about parametrization

Table 3 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.

[Table 3 about here]

Institutional environment. Based on Author (yearb), the size-threshold into formality, l^{ic}, is set at 30 employees[18], the ratio of tax obligations per sale found in the data is $\tau_0 = 0.14$. The effective rate of taxes paid,τ_p, is half the official rate (Gauthier and Goyette, 2014). Audit probabilities for small and larger firms are also taken from the data and set to 0.53 and 0.71 respectively.

Technology. We follow Restuccia and Rogerson (2008) and use $\gamma + \eta = .85$ and attribute parameter values according to capital and labor shares of income,i.e., 1/3 and 2/3, respectively.[19] We assume a rate of depreciation of $\delta = 0.07$.

To assign other technology parameters, we follow Hopenhayn and Rogerson (1993). First, the price of output is normalized to one. Second, we use 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.[20]

[17] See Author (yeara) for a more detailed description of the algorithm.

[18] This 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).

[19] Note that Soderbom and Teal (2004) find a very similar assignment of income shares for Ghana.

[20] Our 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).

The relative demand for labor, equation 25, allows the identification of 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 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 parameter we add as a target in the loss function (26): the share of small firms in the data. This results in an estimate of $n_i = 11$, meaning that initial productivity assumes values over the lowest 11 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:

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

where $0 < \rho < 1$ and ϵ_t is a white noise perturbation. We assume that the distribution of the white noise is a Normal truncated at zero with variance σ_ϵ^2. The AR(1) process is covariance-stationary with $\sigma_z^2 = \sigma_\epsilon^2/(1-\rho^2)$. 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 through a discrete state-space process. This involves constructing a grid for productivity shocks with n_z points and an upper bound given by:

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

A transition matrix $g(z_{t+1}, z_t)$ is then computed using a parameter $p = \frac{1+\rho}{2}$. Given that ρ affects directly the persistence of productivity shocks, we exploit time variation in firms' revenues as observed in the data to pin down this parameter. More particularly, we add as a target in the loss function (26) the first order autocorrelation coefficient of a regression of firms' revenues in 1997 on its two available lags in the data (1996, 1995). We then pin down the value of σ_ϵ^2 using equation 28. Given that, as described above, z_{max} is equal to 3.8, we can use the estimated value for ρ to back out σ_ϵ^2 from equation 28.

Given that changes in the fixed cost of production generate scale effects in models à la Hopenhayn, we estimate c_f by targeting the share of output of large firms in equation (26). The entry-cost is estimated using the free-entry condition 17.

Exogenous firms' exit. For the mortality 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 value of 12% for the mortality rate in Uganda and set φ, the survival rate, to 0.88 . We discuss some implications of this choice below.

Preferences. For discount factors of entrepreneurs in developing countries, Arellano et al. (2012) 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 calculated based on the optimal labor equation 20 for the smallest firms in the model, i.e., those with one employee, the entry level of capital and the lowest shock. The corresponding wage in the data is 1096 USD. 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.[21]

Financial Development. We set the rental rate of capital, r, to 0.21, which is the average commercial bank lending rate for 1997 based on data from the IMF. For the financial development parameter, ϕ, we do not have a direct counterpart in the data. Other papers have used the external finance-to-GDP ratio as a target to calibrate similar parameters (see Buera and Shin,2013). We choose to include the share of credit constrained firms (unable to apply for a loan) among the targets in our simulation-based estimation. In the data, these firms consist mainly of small firms and a few medium firms. We thus make sure in the benchmark that all small firms hit the credit constraint. We expect that variation in this dimension conveys at least some information about the extent of credit market access. Below, we conduct a sensitivity analysis to examine the effect of ϕ on the main variables under study.

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

Summary. We thus estimate four parameters using a simulated method of moments. These parameters are: ρ, ϕ, c_f and the grid point corresponding to the maximum productivity of new entrants, n_i. There are four explicit target moments included in the loss function: (i) the share of output of large firms; (ii) the share of small firms in the size distribution of firms; (iii) the share of small credit constrained firms in the sample; (iv) the first order autocorrelation coefficient of firms' revenues. We also add two additional conditions in the loss function: the requirement that the model reproduces the growth rate of large firms and the requirement that the difference in capital labor ratio between small and medium firms be non-negative. We thus have a u-vector of 4 parameters for ζ and a v-vector of 6 moments.

We use the parametrized model to verify to what extent the input market distortions translate in a size distribution of firms that is highly skewed to the right, patterns in capital-labor ratios and growth rates that are similar to those observed in Uganda.

5.2 Model validation

5.2.1 Model fit

We compare in Table 4 the simulated moments obtained from the calibration with their targets in the data. The model is right on target for the share of small firms and the growth rate of large firms. The model does a reasonable job to replicate the share of output of large firms and the first order autocorrelation coefficient on revenues. The model also captures well the share of small and medium firms being credit constrained. Note that the share of medium firms in the model corresponds approximately to the share of medium firms being credit constrained in the data. However, given the coarse modeling of credit constraints in the model, the share of large firms being credit constrained is largely overestimated. We discuss patterns in capital-labor ratios below.

5.2.2 Out-of-sample restrictions

In this section we examine some model statistics that were not explicitly targeted in the calibration.

[Table 4 about here]

Firm Size Distribution. The model generates an equilibrium size distribution of firms with a 'missing middle' and a thick left tail. Examining the share of firms for the two size bins that were not explicitly targeted we note that the model tends to overestimate the shares of large firms to the expense of medium 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 step-wise jump in audit intensity at the 'formality' threshold for small firms while in reality the increase in audit intensity might not be as stark. However despite these simplifying assumptions, the benchmark produces average sizes that are much in line with those from the data for the whole size distribution but also, more specifically, for small and medium firms. The underestimated average number of employees in large firms is a consequence of the overestimation of the share of large firms in the model.

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 model delivers patterns in growth rates that qualitatively mimic the pattern observed in the data for small and medium firms (see table 4). The magnitudes of the growth rates for small and medium firms are quantitatively much larger than what we estimate for Uganda but note that they lie well within the confidence intervals in the data.

In table 5 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. As can be noted from table 5, the model reproduces a similar pattern as the one observed in the data: growth rates seem to be increasing with size for small and medium firms.

[Table 5 about here]

Capital-labor ratios. We also examine the patterns in capital-labor ratios, using both total or owned capital.[22] Table 4 shows that the average capital-labor ratio, using owned or total capital, is slightly larger for small firms than for medium firms. The ratio for large firms goes

[22]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.

back up and is larger than for medium firms in both simulated data and Ugandan data. These patterns are a reflection of selection, the collateral constraints and the distortion due to the audit probability change.

We can gain more insights into the driving forces lying behind these patterns using a graphical illustration. Figure 2 plots capital-labor ratios using total capital. Notice that each dot represents a cluster of firms with identical productivity. The dots lining up in the left-most part of the graph are mostly new entrants. As mentioned earlier, there are three forces driving the patterns in capital-labor ratios: selection, credit constraints and the tax distortion. In the benchmark calibration, it seems that selection is probably driving the bulk of the result for small and medium firms. Indeed, removing entrants in figure 3, we clearly see that average capital-labor ratios are now increasing across the whole size range. However, this does not imply that the tax distortion has no effect on capital labor ratios in Uganda. We observe in figures 2 and 3 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 4 below, where credit constraints are removed and there is no dispersion within each productivity group. Finally, note that we do not observe as much dispersion for larger firms as for small firms in figures 2 and 3. One potential explanation is that selection plays an important role in discarding sub-optimal decision makers and thus most of the dispersion in the smaller size range is probably due to entrants or young firms which have not yet been forced out of the market. This observation seems consistent with figure 1 where dispersion is decreasing with firm size.

Other non-targeted statistics. The model generates an overall 26% exit rate of which a bit more than half is accounted for by endogenous exit decision. This result is due to the same reason as in Midrigan and Xu (2014) who find that 26% of producers must exit in any given period in order to reproduce the large fraction of young producers in their data. In terms of share of output, the model does a good job at replicating the share of output for small firms. However, the share of output of large firms is overestimated to the expense of the share of medium firms. Finally, note that the second order auto-correlation coefficient on firms' revenues which is not explicitly targeted in the benchmark is close to the one observed in the data.

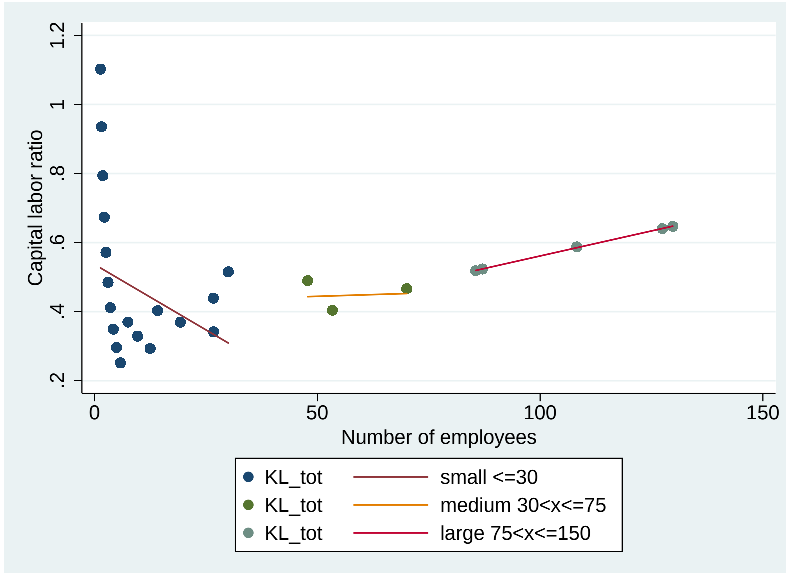

Figure 2: Simulated Capital-Labor Ratios v.s. Firm Size (Total Capital)

5.2.3 Counterfactual Analysis

In this section we examine the qualitative and quantitative implications of the observed distortions within our simple model. We do this by comparing the benchmark equilibrium to 5 alternative scenarios.

We start by examining the first-best allocation. In this scenario, corporate taxes are completely removed and access to credit is unconstrained. This exercise is performed under two different assumptions about the rental rate of capital. First, we use the average observed commercial rates as in the benchmark. Second, we examine what would happen if country-specific risk could be reduced in Uganda. In order to do so, we use data from the World Bank Indicators to run a cross-country regression of lending rates on a measure of institutional quality (strength of legal rights).[23] We then use Uganda's level of measured institutional quality to evaluate its predicted lending rate. We find that if Uganda were to move from its current level in terms of institutional quality to a level comparable to the U.S., one could potentially observe a 25% reduction in the Ugandan lending rate, i.e., from 21% to 15.7% (results available upon request). Note that this corresponds approximately to a one standard deviation in Ugandan lending rates

[23] See `http://data.worldbank.org/indicator/FR.INR.LEND` for lending rates and `http://data.worldbank.org/indicator/IC.LGL.CRED.XQ` for strength of legal rights.

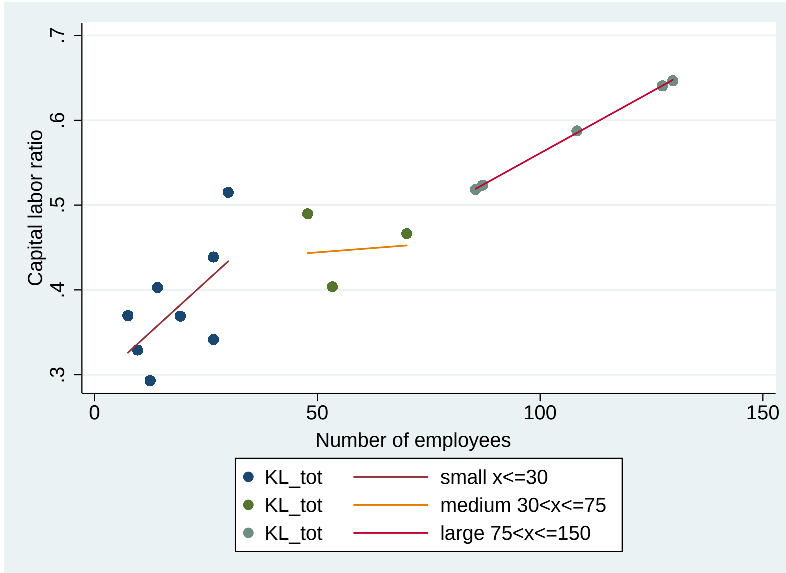

Figure 3: Simulated Capital-Labor Ratios v.s. Firm Size with no entrants (Total Capital)

over the last 30 years.

Next, we assess the relative importance of the different distortions in explaining the efficiency deviation from the first best and run the following experiments: (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 $\phi = \phi^{max} = r/(1-\delta)+1$); (iii) keep the tax distortion and credit constraints, but change the lending rate from its benchmark value, 21%, to the value corresponding to the level of U.S. institutional quality, i.e., 15.7%.

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. Therefore these experiments focus on the effects of pure labor reallocation. Moreover, it is important to note that the adjustment in the wage absorbs part of the productivity gain and decreasing returns to scale also limit the gains from over-accumulation of the factors of production. The results with inelastic labor supply should thus be interpreted as lower bounds on the productivity differences. Of course, it is likely 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 (with constant aggregate labor supply), and that the remaining labor market clearing is borne out by changes in aggregate labor supply. We 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 (results not shown).

Table 6 and 7 summarize the percentage change from the benchmark for key variables for experiments with inelastic and elastic aggregate labor supply, respectively.

[Table 6 about here]

[Table 7 about here]

Experiments 1a and 1b: First-Best In the first-best, corporate taxes are removed and firms can always rent the desired amount of capital. We first examine the results when aggregate labor is kept as in the benchmark in table 6. In experiment 1a, output per worker and aggregate capital increase by 19% and 67% respectively. Given that capital deepening is driving most of the effect on output per worker, we thus report aggregate TFP as a more precise measure of the improvements in the allocation of resources due to the change in the environment. We observe a 3.2% increase in TFP in experiment 1a. The total exit rate decreases by 16%.

Examining a situation where some adjustment in aggregate labor is allowed for (Table 7), we find that output per worker, aggregate capital and TFP increase by 7%, 184% and 51% respectively in experiment 1a. Also, more favorable input prices generate lower selection as total exit is reduced by 29%.

In experiment 1b in tables 6 and 7, we use a rate of rental of capital that could potentially prevail if institutional quality in Uganda was adjusted to the U.S. level, say 15.7%. This brings about the highest changes in aggregate variables. When only the wage is adjusted (table 6), this results in a 34% increase in output per worker. Note that if we let labor and wages adjust partially (table 7), the increase in output per worker is of the order of 26% but exit drops by 37%.

Capital-labor ratios in experiments 1a and 1b are identical across firms, exactly as the theory predicts (see equation 23). The change in the size distribution of firms are such that there are more medium and large firms. Given an unchanged labor supply in table 6, the average number of employees turns out to be lower in all size bins. Finally, no firms face credit constraints in experiments 1a and 1b.

Experiment 2: No corporate taxes From table 6, we note that the wage goes up by 19% in experiment 2. The change in aggregate capital, output per worker and TFP are 20%, 4.4% and -.7% respectively. The drop in TFP is due to the fact that the increase in aggregate capital is much larger than the increase in output per worker. This effect is even more important in table 7 where output per worker is actually decreasing in response to the change in the tax environment. This is partly due to the production function which exhibits decreasing returns to scale.

Examining capital-labor ratios, we see that the selection effect is still present and such that, even after removing the tax distortion, average capital-labor ratios are almost equal for small and medium firms. When we allow for aggregate labor adjustment in table 7, even if a more favorable price of labor reduces selection (exit drops by 29%), capital labor ratios remain higher for small firms compared to medium firms.

However, as seen from the percentage changes in the share of small and medium firms in table 6, several firms (10%) which were small in the benchmark reallocate themselves in the medium size bin (56%) and a few in the large size bin (2.4%). All these firms would be interested in external finance as we observe a similar increase in the share of credit constrained medium and large firms.

Experiment 3: Perfect Capital Markets We now examine the case of unconstrained access to credit, keeping the interest rate at the benchmark level $r = 0.21$. From experiment 3a in table 6, we note that the wage must increase by 15% to keep aggregate labor as in the benchmark.

Aggregate capital, output per worker and TFP go up by 41%, 15% and 3.9%, respectively. Note that the average capital-labor ratio is higher for small firms than for medium firms, because of selection and the regulation distortion. However, we observe constant capital-labor ratios across medium and large firms as the theory predicts. 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.

Lowering the interest rate in conjunction with the removal of credit constraints (experiment 3b in table 6), brings about much larger changes. Output per worker goes up by 29% but the change in TFP remains almost the same due to the large increase in aggregate capital (112%). The effects when we let labor adjust are qualitatively similar but of smaller magnitude (table 7).

In figure 4, we present capital-labor ratios for the perfect credit market case. we clearly see that around 30 employees there is a cluster of firms which substitute capital for labor. Moreover, 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. All dots in figure 4 represent a mass of firms making the same optimal choice of labor and capital.

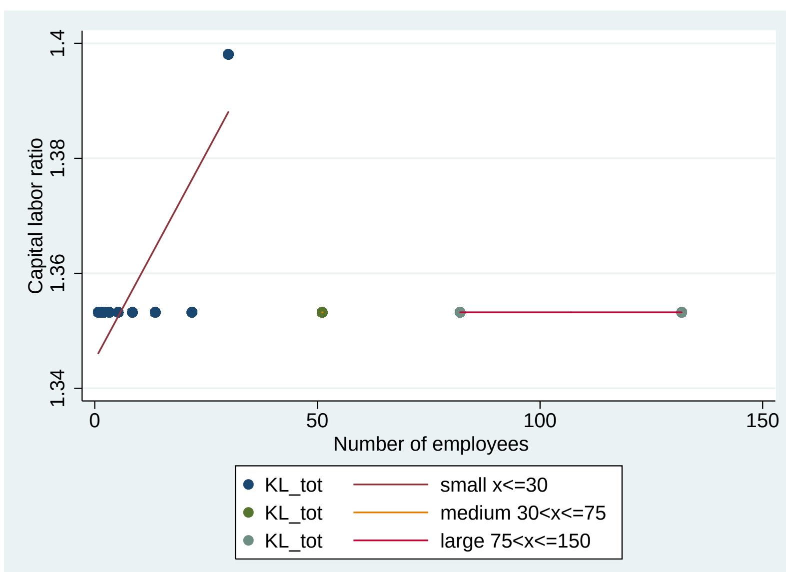

Figure 4: Simulated Capital-Labor Ratios v.s. Firm Size for Experiment 3

Experiment 4: Lowering lending rates We now examine the effect of a change in lending rates from the benchmark value to a value that would potentially prevail if Uganda could straighten up its legal rights to the level of the U.S. In experiment 4 in table 6, we observe an increase in aggregate capital, output per worker and TFP of 11%, 2.7% and -.2% respectively. There are at least two potential explanations for the negligible effect on TFP. First, credit constraint in the benchmark are so important that firms rely mostly on internal finance. As such, they cannot take advantage of a reduction in the rental rate of capital. Second, the effect on capital-labor ratios of a change in the rental rate and the effect of the tax distortion are going in opposite directions. The effects when the wage adjusts are negligible compared to the other experiments and in the same direction as those described in this paragraph as can be inferred from table 7.

Experiment 5: Flat Audit Intensity Flattening the audit intensity yields negligible effects on aggregate labor, capital and output whether the wage is adjusted or not. The overall audit intensity that yields the same tax revenues as in the benchmark is $p_a = 0.72$. As noted in table 6, we observe a reallocation of firms from the small size bin to the medium size bin. The 'missing middle' in the firms' size distribution decreases, suggesting that it is a by-product of the distortions in the tax system and that it is associated to negligibly small losses in aggregate efficiency.

Sensitivity analysis In table 8, we examine the response of the main aggregate variables to changes in ϕ when aggregate labor is kept as in the benchmark. The direction of the changes in output per worker, TFP and aggregate capital are as expected. Also, changes in ϕ do not affect exit at any time.

As credit constraints are loosened (higher ϕ), aggregate rented capital increases while aggregate owned capital decreases. This should come as no surprise: entrepreneurs prefer to make optimal capital rental choices after observing their current productivity shock rather than accumulate potentially idle capital. Also, note that capital labor ratios for medium and large firms are converging. However, capital-labor ratios of small and medium firms change non-linearly with the level of credit constraints. There are at least 2 potential explanations for this result. First, this non-linear relationship is partly due to the reallocation of firms across size bins and changes in the average number of employees per size bin. Second, the unstable pattern in capital-labor ratios may be a consequence of the coarse approximation of credit constraints in the model.

Examining how firms reallocate themselves across size bins, we note that, if credit constraints have more bite (lower ϕ) then: 1) firms are shifting towards the left tail of the size distribution, 2) there are fewer large firms, 3) these large firms are now bigger on average. Hence, credit constraints play an important role in shaping the overall size distribution of firms by exacerbating skewness.

Finally, note that for ϕ between 0 and 0.2 changes in the share of credit constrained firms match the exact changes in the share of firms across size bins. However, for ϕ between 0.6 to 1, the change in the share of small firms being credit constrained is larger in magnitude than the change in the share of small firms, implying that less small firms are credit constrained.

Discussion It is worth stressing again that the adjustment of the wage, as well as the high rental rate on capital, drive down the overall efficiency gains in each experiments and these results should be considered as lower bounds of the likely effects.

Using table 6, the model shows that removing the tax distortion (experiment 2) would reduce the gap in output per worker between the benchmark and a first-best scenario with r=0.157 (experiment 1b) by 13%. The introduction of perfect capital markets (experiment 3a) would further reduce this gap by 44% while a simple change in the rental rate of capital would only reduce the gap by 8% (experiment 4). However, it really is the interaction between the cost of capital and the removal of credit constraints (experiment 3b) that goes the longest way in explaining the difference in output per worker between the benchmark and the first-best as it accounts for 85% of the gap.

Also, the results from experiment 5 suggest that the 'missing middle' in the size distribution of firms is due to the tax distortion, rather than imperfect credit markets. The 'missing middle' in the firms' size distribution is a by-product of the distortion in the tax system and is associated to relatively small losses in aggregate efficiency. Our results suggest that the 'missing middle' simply reflects the optimal responses of forward looking firms to the tax environment, and that such responses are quite effective in reallocating resources so to minimize the impact of the distortion. However, the sensitivity analysis shows that credit constraints play a larger role in shaping the whole size distribution of firms and that high levels of financial underdevelopment are associated with thicker left tails and higher levels of asymmetry in the size distribution of firms.

6 Conclusion

We analyze the allocation and efficiency implications of two large and observable input market distortions, due to uneven audit intensity and limited access to firm's finance. We find that amending a growth model of heterogeneous firms to account for these distortions generates an equilibrium firms' size distribution, and patterns of capital-labor ratios and growth, which are qualitatively similar to those observed in micro-data for Uganda. The numerical counterpart of the model fits the data reasonably well in several dimensions and is used to assess the implications of each distortion in turn, through counterfactual experiments.

We find that output per worker in the benchmark economy is 19 percent below the first-best level, when we keep the rental rate of capital constant and equal to the one observed in the data. However the difference is larger (34%) when the rental rate of capital in the first-best is set equal to 15.7%, corresponding to a situation where Ugandan institutional quality would be matched to the U.S. level. This large discrepancy suggests that credit market imperfections play a large role in the sub-optimal allocation of resources.

Removing the tax distortion would close the gap in output per worker by 13%. However, introducing perfect capital markets would reduce this gap furthermore by 44%. However, removing credit constraints and using a rate of rental adjusted for the level of institutional quality observed in developed economies would generate much larger gains and account for a 85% reduction in the observed gap.

One key feature of the model is that the discrete jump in audit probability and total tax liabilities generates a cluster of firms below the exogenous size-threshold. Such a phenomenon is not specific to the environment at hand (see for instance Hsieh and Olken, 2014). As mentioned in the introduction, many developing countries have adopted a tax system with a VAT structure like the one prevailing in Uganda. Moreover, similar clusters could arise wherever a sharp change in policy conflicts with agents' incentives. However, our results indicate that the 'missing middle' simply reflects the optimal responses of forward looking firms to the tax environment and corresponds to small efficiency losses. However, the main result of the paper is that financial underdevelopment plays a major role in shaping size distribution of firms with no less important consequences on output per worker. Paraphrasing Restuccia and Rogerson (2008), it seems that the research agenda should now aim at identifying and quantifying the specific components of the financial system which generate resource misallocation.

Acknowledgements:

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

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. NBER WP 14060.

Arellano, C., Bai, Y., Zhang, J., 2012. Firm dynamics and financial development. Journal of Monetary Economics 59 (6), 533–549.

Author, Unpublished resultsa. Other Paper 1 by author. Journal V, pages.

Author, Unpublished resultsb. Other Paper 3 by author. Journal V, pages.

Author, yeara. Doctoral disseration. Thesis V, pages.

Author, yearb. Other Paper 2 by author. Journal V, pages.

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.

Buera, F. J., Kaboski, J. P., Shin, Y., 2011. Finance and development: A tale of two sectors. American Economic Review 101 (5), 1964–2002.

Buera, F. J., Shin, Y., 2013. Financial frictions and the persistence of history: A quantitative exploration. Journal of Political Economy 121 (2), 221–272.

Cihak, M., Demirguc-Kunt, A., Feyen, E., Levine, R., 2013. Financial development in 205 economies, 1960 to 2010. Journal of Financial Perspectives 1 (2), 17–36.

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., 2014. Small Scale Reservation Laws and the Missallocation of Production Factors. Journal of Monetary Economics 66, 193–209.

Gauthier, B., Goyette, J., 2014. Taxation and corruption: theory and firm-level evidence from uganda. Applied Economics 46 (23), 2755–65.

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.

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.-T., Klenow, P. J., 2009. Misallocation and manufacturing tfp in china and india. Quarterly Journal of Economics 124 (4).

Hsieh, C.-T., Olken, B. A., 2014. The missing missing middle. The Journal of Economic Perspectives 28 (3), 89–108.

Keen, M., Lockwood, B., 2010. The value added tax: its causes and consequences. Journal of Development Economics 92, 138–151.

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.

Midrigan, V., Xu, D. Y., 2014. Finance and misallocation: Evidence from plant-level data. American Economic Review 104 (2), 422–458.

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.

Rosenzweig, M. R., Wolpin, K. I., 1993. Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in india. Journal of political economy, 223–244.

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., 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), 11–44.

Table 1: 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: 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.

Table 3: 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	0.39	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	0.96	Persistence	estimated using model
σ_ϵ^2	0.06	Variance White Noise	estimated using model
w	1096$/yr	Wage	measured from data
φ	0.88	Firms' survival rate	assigned, external estimate
n_i	11	Maximum initial productivity	estimated using model
c_f	0.5	Fixed cost of production	estimated using model
c_e	22.5	Entry cost	Free-entry condition
Preferences			
β	0.97	Intertemporal discount factor	assigned, external estimate
D	1.22	Disutility of labor	estimate using model

Table 4: Targets and non-targeted Statistics

Variable	Firm's size	Benchmark	Data	
		Average	Average	Std. Err.
Share of firms	Small (target)	.48	.48	
	Medium	.06	.20	
	Large	.46	.32	
Average size (employees)	All firms	131	124	259
	Small	10	14	8
	Medium	57	50	13
	Large	267	334	379
Output growth	All firms	.27	.08	.54
	Small	.31	.02	.43
	Medium	.69	.13	.59
	Large (target)	.17	.17	.67
Share of output	Small	.04	.03	
	Medium	.03	.14	
	Large (target)	.93	.84	
Owned capital per worker	Small	28.4		
	Medium	27.6		
	Large	57.3		
Total capital per worker	Small (target)	41.5	78	129
	Medium	40.3	38	35
	Large	83.8	50	97
Exit	Total	26%		
	Exogenous	12%		
	Endogenous	14%		
AR(2)	Revenues t-1 (target)	1.00	1.05	.07
	Revenues t-2	-.07	-.05	.07
Share of credit constrained firms	Small (target)	1.0	1.0	
	Medium	1.0	0.05	
	Large	1.0	0.0	

Table 5: Revenue growth and Number of Employees: OLS Estimates

Dependent Variable: Growth rate of Revenues (1996-1997)[a]		
	Data	Benchmark
# of Employees	.005***	.007
	(.002)	(.001)
Constant	-.06	.18
	(.06)	(.03)
R^2	.06	.34
# of observations	157	149
Note: *=10%, **=5%,***=1%; Std. dev. in parenthesis		

Table 6: Experiments with inelastic aggregate labor supply

Experiment	1a.First-Best	1b. First-Best	2. No corporate tax	3a. Unrestricted access to credit	3b. Unrestricted access to credit	4. Change in rental rate	5. Flat audit intensity
	r=0.21	r=0.157	r=0.21	r=0.21	r=0.157	r=0.157	r=0.21
Change in:				(% change from the benchmark)			
wage	36	53	19	15	29	2.7	-1.4
aggregate output	19	34	4.4	15	29	2.7	-1.2
aggregate TFP	3.2	3.2	-.7	3.9	4.1	-.2	-1.0
aggregate capital	67	151	20	41	112	11	-1.0
Ratio capital/worker				Ratio			
small/medium firms	1.0	1.0	.99	1.01	1.01	1.03	1.08
medium/large firms	1.0	1.0	.48	1.0	1.0	.47	.47
Change in share of				(% change from the benchmark)			
small firms	-62	-62	-10	-47	-48	0	-4.9
medium firms	145	145	56	37	42	0	37
large firms	44	44	2.4	43	44	0	-.1
Change nb of workers				(% change from the benchmark)			
small firms	-12	-12	-19	36	52	-3.6	-14
medium firms	-24	-24	-20	-10	-11	-4.5	-12
large firms	-31	-31	2.1	-30	-31	.3	.3
Credit constraints				(% change from the benchmark)			
small firms	-100	-100	-10	-100	-100	0	-4.8
medium firms	-100	-100	56	-100	-100	0	37
large firms	-100	-100	2.4	-100	-100	0	-.1
Exit				(% change from the benchmark)			
total	-16	-16	-16	0	-16	0	0
endogenous	-40	-40	-40	0	-40	0	0
exogenous	0	0	0	0	0	0	0

Table 7: Experiments with aggregate labor adjustment

Experiment	1a. First-Best r=0.21	1b. First-Best r=0.157	2. No corporate tax r=0.21	3a. Unrestricted access to credit r=0.21	3b. Unrestricted access to credit r=0.157	4. Change in rental rate r=0.157
Change in:			(% change from the benchmark)			
wage	18	26.5	9.5	7.5	14.5	1.35
aggregate output	103	165	42	46	98	7
aggregate TFP	51	68	24	24	42	3
aggregate capital	184	396	62	80	226	15
aggregate labor	96	139	48	36	73	5
Ratio capital/worker			Ratio			
small/medium firms	1.0	1.0	1.02	1.03	1.02	1.03
medium/large firms	1.0	1.0	.46	1.0	1.0	.46
Change in share of			(% change from the benchmark)			
small firms	-72	-79	-19	-48	-61	0
medium firms	90	52	69	42	11	0
large firms	62	75	10	44	62	0
Change nb.of workers			(% change from the benchmark)			
small firms	21	6	-8	77	68	-.8
medium firms	-6	-27	-17	21	-3	-.8
large firms	25	44	37	-5	11	6
Credit constraints			(% change from the benchmark)			
small firms	-100	-100	-19	-100	-100	0
medium firms	-100	-100	69	-100	-100	0
large firms	-100	-100	10	-100	-100	0
Exit			(% change from the benchmark)			
total	-29	-37	-29	-17	-29	0
exogenous	0	0	0	0	0	0
endogenous	-53	-69	-53	-31	-53	0

Table 8: Sensitivity analysis of financial development parameter

Experiment	$\phi = 0$	$\phi = 0.2$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 1.0$
Change in:	(% change from the benchmark)				
wage	-5	-2.5	4	7	11
aggregate output	-5	-2.1	4	7	11
aggregate TFP	-1.5	-.6	1	2	3
aggregate capital total	-12	-5	9	18	29
aggregate capital rented	-100	-51	69	143	234
aggregate capital owned	29	16	-19	-40	-65
Ratio capital per employee	Ratio				
small/medium firms	.98	1.02	1.02	1.02	.89
medium/large firms	.45	.45	.53	.61	.83
Change in share of	(% change from the benchmark)				
small firms	10	4	-5	-16	-21
medium firms	-10	31	6	14	-29
large firms	-8	-8	4	14	26
Change in number of employees	(% change from the benchmark)				
small firms	3	-2	4	-1	22
medium firms	7	8	.9	-12	6
large firms	9	8	-4	-12	-20
Credit constraints	(% change from the benchmark)				
small firms	10	4	-9	-24	-38
medium firms	-10	31	6	14	-29
large firms	-8	-8	4	14	26
Exit	(% change from the benchmark)				
total	0	0	0	0	0
exogenous	0	0	0	0	0
endogenous	0	0	0	0	0