Firm Size Dynamics in the Aggregate Economy*

Esteban Rossi-Hansberg Mark L.J. Wright

Stanford University Stanford University

July 10, 2004

Abstract

Why do firm growth rates decline with size? What determines the size distribution of firms? This paper addresses these questions in a dynamic model of firm size with entry and exit that emphasizes the accumulation of industry specific factors in response to industry specific productivity shocks. The emphasis on the accumulation and allocation of specific factors leads to new implications for the relationship between capital intensity, firm size, and firm dynamics. We show that these relationships explain a large part of the sectoral heterogeneity in firm sizes and dynamics observed in the US.

INTRODUCTION

Firm sizes dynamics are *scale dependent*: small firms grow faster than large firms, and exit rates decline with size. Scale dependence in growth and exit rates is also systematically reflected in the size distribution of firms. All of these facts have been

*We thank Liran Einav, Bob Hall, Pete Klenow, Narayana Kocherlakota and numerous seminar participants for helpful comments, Trey Cole of the US Census Bureau for his help in constructing the database, and Adam Cagliarini for outstanding research assistance.

documented over different time periods, industries, and countries, which is surprising given the enormous diversity of institutions, market structures, and technology. Moreover, the robustness of these facts demands a theory that emphasize forces common to a variety of circumstances. In this paper we propose a theory that relies on such a force: the response of production decisions to the allocation and accumulation of industry specific factors. Differences in the importance of industry specific factors across sectors then lead to cross-sectoral variation in the degree of scale dependence within a sector. We present evidence from a new dataset to document these facts for the US economy. We find that, as implied by our theory, differences in the intensities of specific factors across sectors are related to significant differences in the degree of scale dependence in firms dynamics.

A large literature beginning with Gibrat (1931) examined the size distribution of firms. The observed size distribution is sensitive to the definition of a firm. Figure 1 presents the densities of establishments (single unit plants) and enterprises (operations under common ownership or control) for the US economy in 2000 and compares them to a commonly used benchmark: a Pareto distribution with shape coefficient one (see, for example, Axtell (2001)). The figure shows that, although the enterprise and establishment size distributions are similar, reflecting the fact that only the very large enterprises possess more than a single establishment, they differ substantially for large size classes. Both distributions have thinner tails than the benchmark. If production units are distributed according to a Pareto distribution, the natural logarithm of the share of production units greater than a particular employment size varies linearly with the natural logarithm of employment. If the Pareto distribution has a shape coefficient of one, the slope of the line is minus one. If, however, the tails of the actual distribution are thinner than the tails of a Pareto distribution, as in Figure 1, the relationship is concave and not linear. In Figure 2, one can see that, even though the Pareto distribution is a good approximation for large enterprises, the curves are clearly concave, especially for establishments. The theory we develop below refers to the technology of a single production unit and does not address questions of ownership or control. Therefore, throughout the paper we define firms as establishments and focus solely on them.

Figure 1:
Density Function of Establishments and Enterprises in 2000

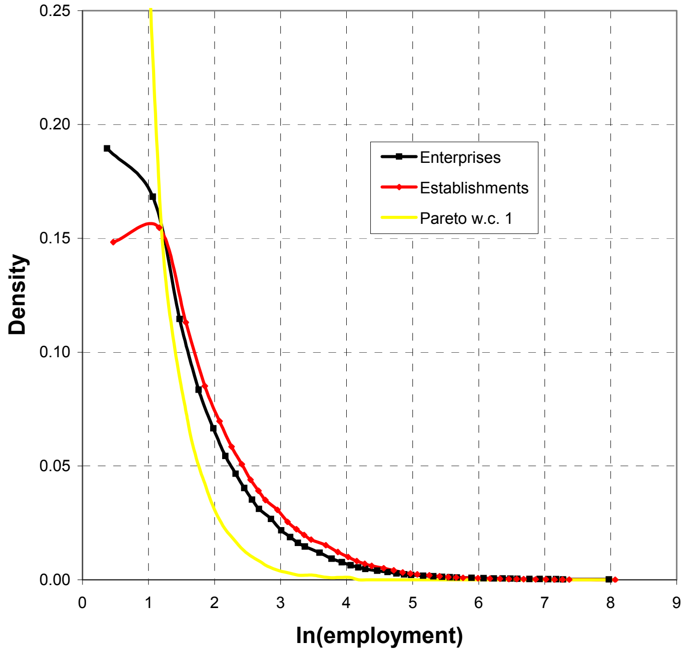
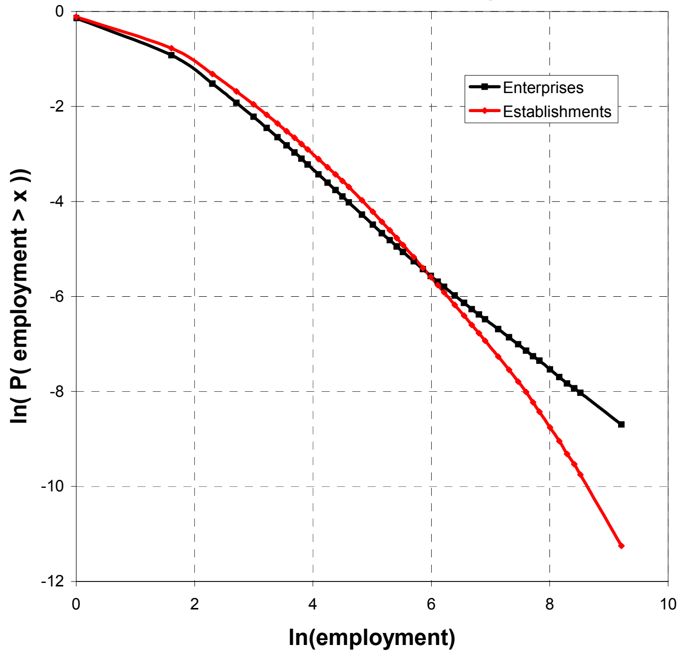


Figure 2:
Distribution of Establishments and Enterprises Sizes in 2000

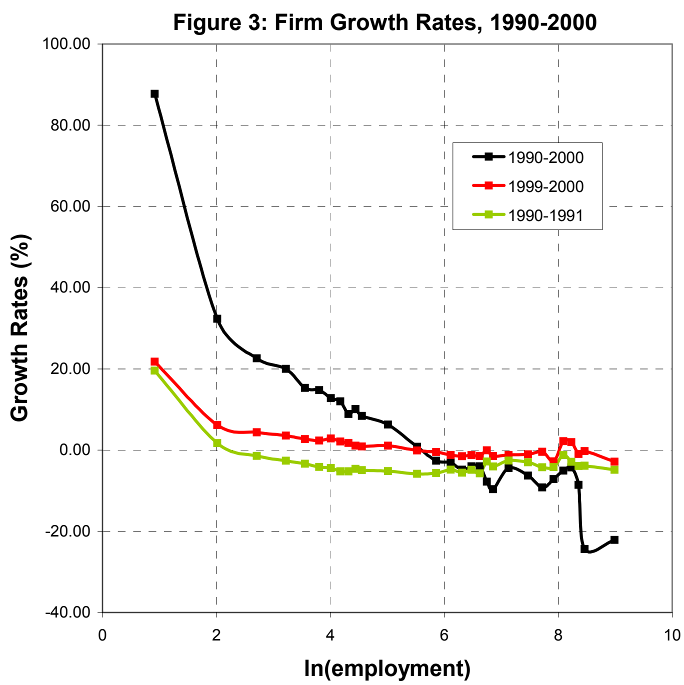


The firm size distribution reflects the dynamics of firm sizes in the economy. Looking at firm growth rates, while many authors agree with the conclusion of Scherer (1980) that scale independent growth "is not a bad first approximation", it is clear that it is only an approximation and that some of the approximation errors are systematic.[1] Perhaps the best established of these is that small firms grow faster than large firms, at least when attention is restricted to those firms that remain in operation.[2] This is illustrated in Figure 3 which plots growth rates by firm size for the US

[1] See, for example, the surveys by Geroski 1995, Sutton 1997, and Caves 1998, who also document the robustness of these results across time, industries and countries.

[2] This fact was most forcefully demonstrated by Mansfield (1962) in his study of firms in the steel,

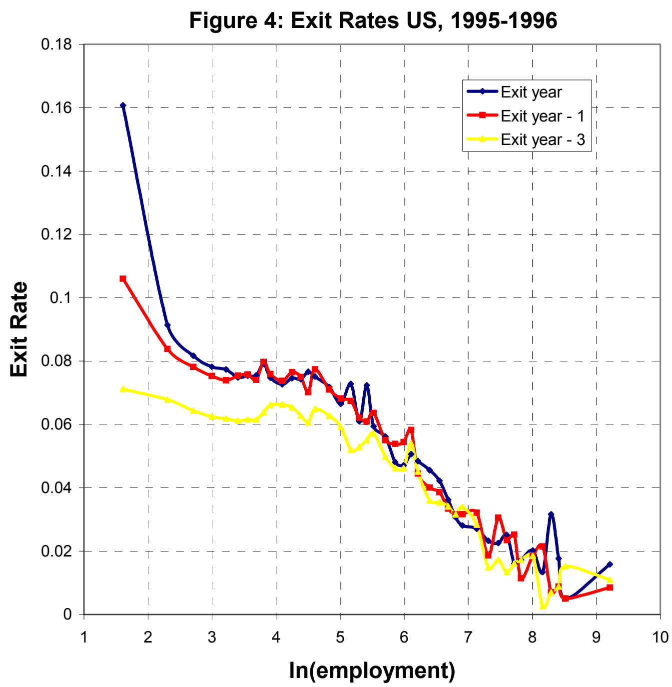
over both one and ten year intervals. This figure shows that the difference in growth rates between small and large firms can be as large as twenty per-cent within a year, and that the accumulated effect of this pattern over a decade leads to differences of more than one-hundred per-cent between small and large firms. Moreover, this scale dependence in growth rates is not limited to the smallest firms, and is significant throughout the size distribution.



In a typical period, a substantial fraction of production units turn over: some units exit, while new ones are created. Mansfield (1962) was one of the first to emphasize the importance of turnover and to find that smaller firms were more likely to exit. This scale dependence in exit rates is illustrated in Figure 4 which follows the cohort of firms that exited between 1995 and 1996 in the years leading up to their death.

petroleum, tire and automobile industries. More recent work by Hall (1987) and Evans (1987a,b) using data on firms, and by Dunne, Roberts and Samuelson (1989a,b) on manufacturing plants, has confirmed this finding.

Several features in this figure should be noted. First, exit rates decline substantially with size. Second, only the smallest firms decline in size in the years leading up to their death. In fact, this graph shows that there is no significant decline within firms with more than 150 employees in the three years leading up to their exit.



To address these facts, we propose a theory of firm dynamics based on the accumulation of industry specific factors. We present a stochastic growth model with multiple goods. The set of goods in the economy is divided into sub-groups that we call sectors. Each sector is in turn formed by a collection of goods that we call industries (think of 2 or 3 digit NAICS codes for sectors, and 4 or more for industries). Firms operate in only one industry and hire labor and an industry specific factor. As long as technology exhibits diminishing returns to the specific factor at the firm level, and this is preserved by aggregation within an industry, an abundance of this factor leads to low rates of return and, therefore, consumers invest relatively small amounts in this industry. Conversely, if the stock of the specific factor is relatively low, rates of return are high and consumers invest heavily in this industry. This process, which is at the heart of the resource allocation mechanism in the economy, leads to mean reversion in the stocks of industry specific factors. As long as firms respond monotonically to fluctuations in factor prices driven by the stock of specific factors, mean reversion in these stocks leads to mean reversion in firm sizes. This results in small firms growing faster than large firms.

Given the level of employment in the industry, increases in average firm sizes imply that some firms exit. The extent to which employment in the industry varies depends on the degree of substitutability in consumption determined by preferences. As long as the degree of substitutability is not too large, employment at the industry level does not increase enough to offset the larger firm sizes, and firms exit. Since small firms grow faster than large firms, the exit rate is largest for small firms: scale dependence in exit rates. We can then combine the implications of the model for growth and exit to show that in the long run the distribution of firm sizes in a sector converges to an invariant distribution that has thinner tails than the Pareto distribution with coefficient one.

The driving force behind all of these results is the accumulation of industry specific factors. As a result, the mechanism is robust to a variety of different environments. To establish this, we also consider different production technologies (such as the specification of costs and factor shares), within industry firm heterogeneity, differences in institutions (for example, tax policy), and differences in the form of competition.

Sectors vary widely in their factor intensities. This is related to the degree of diminishing returns in each of these factors (the decline in the marginal product of this factor with respect to its quantity), and hence implies that the strength of our proposed mechanism should differ across sectors. For example, the extent to which growth rates decline with firm size within a sector varies with the share of the specific factor in that sector. To assess this implication of our theory, we focus on physical capital as the specific factor in the data, given that linearity in the accumulation of industry specific human capital prevents the rate of accumulation of this factor from varying with factor stocks. Using a new dataset commissioned from the US Census Bureau on firm growth and exit rates, as well as firm size distributions, for very fine size categories and 2 digit SIC industries, we test for the relationship between capital shares and firm scale dependence. We first test the implication on growth rates and show that, as predicted by the theory, there is a positive and significant relationship between scale dependence in growth rates and the degree of diminishing returns as parameterized by capital shares. We then proceed to show that this same relationship is reflected in significant differences in the size distribution of firms across sectors. These differences are also large: in order to make the size distribution of firms in the capital intensive manufacturing industry conform to the size distribution of firms in the labor intensive educational services sector, we would need to take roughly three million employees (about twenty per-cent of total manufacturing employment) from medium size manufacturing firms (between 50 and 1000 employees), and reallocate two million to very large, and one million to very small, firms. To the best of our knowledge, this is the first study to make use of detailed firm size data for the entire non-farm private sector. This allows us to uncover these novel empirical regularities predicted by our theory.[3]

Most recent theoretical attempts to explain the size distribution of firms have focused on the dynamics of firms in an industry assuming elastically supplied factors of

[3] Relatively little work has examined cross-industry differences in firm sizes. In terms of firm growth rates, Audretsch et al (2002) found that Gibrat's Law is a better approximation for the Dutch services sector than it is for the manufacturing sector. In terms of entry and exit, Geroski (1983) found that gross entry and exit rates of firms are positively correlated across industries, while Geroski and Schwalbach (1991) found that turnover rankings were common across countries. Orr (1974), Gorecki (1976), Hause and Du Rietz (1984) and MacDonald (1986) all found that firm exit rates were negatively related to measures of capital intensity by industry.

production. These frameworks have proven useful in understanding the dynamics and distribution of firms within an industry. Our mechanism operates at a more aggregate level in which the identities of individual firms within an industry are irrelevant and focuses on changes in relative factor prices across industries. One can imagine future extensions of this work that combine both approaches. Another characteristic of most of these frameworks is that they generate scale dependence via selection mechanisms: unsuccessful firms decline and exit. In Jovanovic (1982), this selection occurs as firms learn about their productivity, while in Hopenhayn (1992) and Ericson and Pakes (1995) a sequence of bad productivity shocks leads firms to exit. In Kortum and Klette (2003), it occurs as firms add and subtract product lines in response to their own and competitors' investments in research and development. We acknowledge that these type of effects may be important for small firms, but we believe that they may be less relevant for the scale dependence observed across medium sized and large firms.

Another mechanism that has its main impact on small firms is the presence of imperfections in financial markets as in Cabral and Mata (2003), Clementi and Hopenhayn (2002), Albuquerque and Hopenhayn (2002) and Cooley and Quadrini (2001). Cabral and Mata (2003) present evidence that the size distribution of a cohort of surviving firms shifts to the right and approaches a log-normal distribution over time. They read this as support for the existence of financial constraints on small firms. However, our model is also consistent with this finding. Since small firms grow faster than large firms, and enter more in absolute terms, following a cohort of surviving firms over time results in distributions where the mass of firms shifts to the right. As emphasized by Cooley and Quadrini (2001), both age and size effects are independently important, and we focus mostly on the latter. Other models, for example Lucas (1978) and Garicano and Rossi-Hansberg (2004), produce a size distribution for firms that inherits the properties of the distribution of managerial ability in the population.

In contrast to all of these mechanisms, our model focuses upon the specificity of factors of production in an industry. None of the other theories have this element. In common with Kortum and Klette (2003), our theory has the advantage of simultaneously producing endogenous growth rates, size distributions, and exit rates that can be studied analytically. Many of the mechanisms in the literature undoubtedly contribute towards an explanation of these facts. This paper shows, we believe, that the accumulation of specific factors matters too.

The rest of this paper is structured as follows. Section 2 develops our theory in detail for the case in which firms act competitively and derives the key empirical predictions of our theory. A number of extensions, designed to show the robustness of our mechanism and its predictions to changes in the institutional environment, are presented in Section 3. Section 4 describes our data, and presents results that show that firm growth rates and the firm size distribution vary with observable industry characteristics in precisely the way predicted by our theory. Section 5 concludes.

THE MODEL

We present a stochastic dynamic aggregate model in which firms are perfectly competitive. Labor is mobile across all industries, while there exists a distinct capital good specific to each industry. The model of the firm is standard: fixed costs plus increasing marginal costs of production imply a U-shaped average cost curve, while free entry and exit of firms ensures that all firms in an industry operate at the bottom of their average cost curves. As the focus is upon the allocation of factors across firms and industries, the demand side of the model is kept as simple as possible by assuming logarithmic preferences. This assumption, combined with Cobb-Douglas production functions and log-linear depreciation, ensures that we are able to solve the entire model in closed form.

Households

The economy is populated by a unit measure of identical small households. At the beginning of time, the household has N_0 members, and over time the number of members of the household N_t grows exogenously at rate g_N. Households order their preferences over state contingent streams of many distinct consumption goods and do not value leisure. In describing this economy it is important to distinguish between these goods by what we refer to as a *sector* and an *industry*. We assume that there are S sectors in this economy, and that each sector contains J_s industries, where $s = 1, ..., S$. Each industry produces a single distinct good so that there are $J = \Sigma_{s=1}^{S} J_s$ goods being produced in this economy. In thinking about the data, we define our sectors to be roughly comparable to the list of 3 digit NAICS industries, while our industries map into NAICS industries at a finer level of disaggregation.

Households order their preferences over state contingent streams of each of the J distinct consumption goods $\{C_{tj}\}_{j=1}^{J}$ according to

$$(1-\delta)E_0\left[\sum_{t=0}^{\infty}\delta^t N_t\left(\sum_{j=1}^{J}\theta_j \ln\left(\frac{C_{tj}}{N_t}\right)\right)\right], \tag{1}$$

where δ is the discount factor of the household, θ_j is the consumption share of good j, and E_0 is an expectation operator conditioned on information available to the household at the beginning of time. This function reflects the fact that at any point in time, each of the N_t members of the household consumes an equal share of the households total bundles of consumption, and that the household as a whole sums the valuations of each of it's members. We assume that each industry within a sector has the same consumption share: $\theta_j = \theta_i$ for all i, j in sector s.

In each period, each member of the household is endowed with one unit of time which the household can allocate to work in any one of the J industries, so that if we denote by N_{tj} the amount of time worked in industry j, we have

$$\sum_{j=1}^{J} N_{tj} \leq N_t. \tag{2}$$

Households also rent out their stocks of each of the J industry-specific factors, which we refer to as capital, K_{tj}. Letting X_{tj} denote investment in the industry j capital good, which is assumed to be in terms of the deferred output of the industry, capital accumulates according to the log-linear form

$$K_{t+1j} = K_{tj}^{\omega_j} X_{tj}^{1-\omega_j}. \tag{3}$$

Here ω_j captures the importance of past capital stocks to the amount of capital next period: if ω_j is one, capital does not evolve and is a fixed factor; if ω_j is zero, capital depreciates fully each period. We also assume that this parameter is identical across all industries within a sector; that is, $\omega_j = \omega_i$ for all i, j in sector s. This is one point at which our distinction between sectors and industries becomes relevant: although the stocks of the specific factors may differ across industries, these factors are accumulated via the same investment technology. The household begins with initial stocks of these specific factors denoted by K_{0j}.

Firms

Production within each industry takes place in production units that we call firms. Each firm in industry j at time t has access to the same production technology and hires labor n_{tj} and industry-j-specific capital k_{tj} to produce output according to

$$y_{tj} = A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j}, \tag{4}$$

where A_{tj} is an industry specific productivity shock that obeys

$$\ln A_{t+1j} = \rho_j \ln A_{tj} + \varepsilon_{tj},$$

for some $0 \leq \rho_j \leq 1$, and where ε_t is a random i.i.d. shock with c.d.f. G. It is assumed that the input elasticities sum to less than one ($\alpha_j + \beta_j < 1$) so that firms face a technology with decreasing returns to scale, and that input elasticities and the autocorrelation in productivity are common across all industries within a sector: $\alpha_j = \alpha_i$, $\beta_j = \beta_i$ and $\rho_j = \rho_i$ for all i, j in sector s.

The costs of the firm are given by the cost of inputs, that can be rented at prices r_{tj} for physical capital, and w_{tj} for labor, all of which are normalized by the price of the good being produced, and a fixed cost paid every period in which the firm is active, F_j. Then the problem of the firm is to maximize profits

$$\max_{k_{tj}, n_{tj}} \Pi \equiv \max_{k_{tj}, n_{tj}} A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j} - r_{tj} k_{tj} - w_{tj} n_{tj} - F_j, \tag{5}$$

which gives rise to the usual first order conditions

$$\alpha_j \frac{y_{tj}}{k_{tj}} = r_{tj}, \ \beta_j \frac{y_{tj}}{n_{tj}} = w_{tj}.$$

Firms can freely enter the industry at any point in time so that profits must be zero in equilibrium, or

$$A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j} = r_{tj} k_{tj} + w_{tj} n_{tj} + F_j.$$

Using zero-profits and the first order conditions above, we obtain

$$y_{tj} = \frac{F_j}{1 - \alpha_j - \beta_j}. \tag{6}$$

This implies that if fixed costs and technology parameters are constant, firm output is also constant. This is the result of the assumption that the production function is Cobb-Douglas, and that fixed costs are denominated in terms of the good being produced. We relax this assumption in Section 3.

Let N_{tj}, K_{tj} denote the labor used and physical capital available for production in industry j at time t, and let μ_{tj} denote the number of firms producing in industry

j. Since all firms producing in industry j are identical, equilibrium in factor markets implies that

$$k_{tj} = \frac{K_{tj}}{\mu_{tj}} \text{ and } n_{tj} = \frac{N_{tj}}{\mu_{tj}}.$$

Then,

$$y_{tj} = A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j} = A_{tj} \left(\frac{K_{tj}}{\mu_{tj}} \right)^{\alpha_j} \left(\frac{N_{tj}}{\mu_{tj}} \right)^{\beta_j}. \tag{7}$$

Combining (6) and (7) we obtain an expression for the number of firms in an industry,

$$\mu_{tj} = \left[\frac{A_{tj} \left(1 - \alpha_j - \beta_j\right)}{F_j} \right]^{\frac{1}{\alpha_j + \beta_j}} K_{tj}^{\hat{\alpha}_j} N_{tj}^{1-\hat{\alpha}_j}, \tag{8}$$

where

$$\hat{\alpha}_j = \frac{\alpha_j}{\alpha_j + \beta_j},$$

is the share of capital in factor costs. Hence, total industry output at time t, Y_{tj}, is given by

$$Y_{tj} = \mu_{tj} y_{tj} = \left[\frac{1 - \alpha_j - \beta_j}{F_j} \right]^{\frac{1-\alpha_j-\beta_j}{\alpha_j+\beta_j}} A_{tj}^{\frac{1}{\alpha_j+\beta_j}} K_{tj}^{\hat{\alpha}_j} N_{tj}^{1-\hat{\alpha}_j}.$$

This implies that output in the industry is produced using a constant returns to scale technology, with total factor productivity given by

$$TFP_{tj} = \left[\frac{\left(1 - \alpha_j - \beta_j\right)}{F_j} \right]^{\frac{1-\alpha_j-\beta_j}{\alpha_j+\beta_j}} A_{tj}^{\frac{1}{\alpha_j+\beta_j}}. \tag{9}$$

All of these expressions lead to an equilibrium firm size that depends on the amount of factors in the industry and the current productivity shock according to

$$n_{tj} = \frac{N_{tj}}{\mu_{tj}} = \left[\frac{F_j}{A_{tj} \left(1 - \alpha_j - \beta_j\right)} \right]^{\frac{1}{\alpha_j+\beta_j}} K_{tj}^{-\hat{\alpha}_j} N_{tj}^{\hat{\alpha}_j}. \tag{10}$$

Capital accumulation and labor allocation

To complete the characterization of the evolution of firm sizes in this economy, all that is necessary is to characterize the evolution of productivity and factors in equilibrium. If we allow for a non-integer number of firms, the economy satisfies all of the assumptions of the welfare theorems. As we are primarily interested in allocations, and not prices, we proceed by solving the *Social Planning Problem* for this economy: Choose state contingent sequences $\{C_{tj}, X_{tj}, N_{tj}, \mu_{tj}, K_{tj}\}_{t=0,j=1}^{\infty,J}$ so as to maximize

$$(1-\delta)E_0\left[\sum_{t=0}^{\infty}\delta^t N_t\left(\sum_{j=1}^{J}\theta_j \ln\left(\frac{C_{tj}}{N_t}\right)\right)\right], \tag{11}$$

subject to

$$C_{tj} + X_{tj} + F_j\mu_{tj} \leq A_{tj}K_{tj}^{\alpha_j}N_{tj}^{\beta_j}\mu_{tj}^{1-\alpha_j-\beta_j}, \tag{12}$$

and

$$K_{t+1j} = K_{tj}^{\omega_j}X_{tj}^{1-\omega_j}, \tag{13}$$

for all t and j, and

$$N_t = \sum_{j=1}^{J} N_{tj}, \tag{14}$$

for all t.

Inspection of this problem reveals that the choice of the number of firms is entirely static: μ_{tj} only appears in the resource constraint for industry j at time t. This implies that we can first solve for the optimal number of firms before solving for the dynamics of the economy. The first order condition with respect to μ_{tj} is given by

$$F_j = \left(1-\alpha_j-\beta_j\right)A_{tj}K_{tj}^{\alpha_j}N_{tj}^{\beta_j}\mu_{tj}^{-\alpha_j-\beta_j},$$

which implies

$$\mu_{tj} = \left[\frac{A_{tj}\left(1-\alpha_j-\beta_j\right)}{F_j}\right]^{\frac{1}{\alpha_j+\beta_j}} K_{tj}^{\hat{\alpha}_j}N_{tj}^{1-\hat{\alpha}_j}.$$

This is, as expected, exactly the same expression for the number of firms as we derived above from the firms problem in partial equilibrium. Substituting for the optimal number of firms into the resource constraint gives

$$C_{tj} + X_{tj} \leq \hat{A}_{tj} K_{tj}^{\hat{\alpha}_j} N_{tj}^{1-\hat{\alpha}_j} \equiv \left(\alpha_j + \beta_j\right) \left[\frac{1-\alpha_j-\beta_j}{F_j}\right]^{\frac{1-\alpha_j-\beta_j}{\alpha_j+\beta_j}} A_{tj}^{\frac{1}{\alpha_j+\beta_j}} K_{tj}^{\hat{\alpha}_j} N_{tj}^{1-\hat{\alpha}_j}.$$

The result is an entirely standard log-linear multi-sector growth model with a new constant returns to scale production function and transformed technology level $\hat{A}_{tj}$. The solution of this model has the household accumulating a fixed proportion of the output of each industry as industry specific capital

$$X_{tj} = s_j Y_{tj},$$

and allocating a fixed proportion of its labor endowment to work in each industry.

Implications for Firm Growth, Exit, and the Firm Size Distribution

With these results in hand, we can now characterize the evolution of firm sizes in the economy. Taking natural logarithms and differences of the expression for firm size (10) we find that the growth rate of a firm in industry j is given by

$$\ln n_{t+1j} - \ln n_{tj} = \hat{\alpha}_j g_N - \hat{\alpha}_j \left[\ln K_{t+1j} - \ln K_{tj}\right] - \frac{1}{\alpha_j + \beta_j} \left[\ln A_{t+1j} - \ln A_{tj}\right].$$

This implies that a sufficient condition to obtain Gibrat's Law is that productivity shocks are permanent, so that $\ln A_{t+1j} - \ln A_{tj}$ is i.i.d., and that $\hat{\alpha}_j = 0$. This last condition ensures that capital growth disappears from the equation for firm growth and amounts to the assumption that physical capital is not a factor of production. Another sufficient condition is that shocks be permanent and $\omega = 1$ so that capital is a fixed factor. Gibrat's Law can also be obtained as an approximation for the case of temporary shocks and 100% depreciation ($\omega = 0$) for α_j close to 1. However, in the case that $\alpha_j = 1$, the marginal costs of firms are constant so firms would like to be as large as possible. In this case, there is only one firm per industry, the size of the firm is the same as the size of the industry, and the firm grows at a constant rate

g_N. In this case the size distribution of firms is not stochastic, and any distribution of workers across industries is an invariant distribution.

Using the accumulation equation for industry specific capital and the form of investment derived above, we can recursively substitute for the growth of capital to obtain the long run growth rate of firms

$$\begin{aligned}\ln n_{t+1j} - \ln n_{tj} &= -\frac{1}{\alpha_j + \beta_j}\left\{\ln A_{t+1j} + \left(\left(1-\omega_j\right)\hat{\alpha}_j - 1\right)\ln(A_{tj})\right\} \\ &\quad + \frac{\left(1-\omega_j\right)\hat{\alpha}_j}{\alpha_j + \beta_j}\sum_{s=1}^{\infty}\left(1-\omega_j\right)\left(1-\hat{\alpha}_j\right)\left(\omega_j + \left(1-\omega_j\right)\hat{\alpha}_j\right)^{s-1}\ln(A_{t-sj}).\end{aligned} \tag{15}$$

This equation shows that the effect of productivity shocks on firm growth is shaped by the accumulation and utilization of capital in production. When an industry has received a relatively large sequence of shocks in the past, capital is abundant and the optimal firm size is small as firms substitute from using labor to capital. However, as rental rates are low, the incentive to accumulate capital is weak and over time the capital stock falls and firm sizes rise. The rate at which reversion to the mean occurs depends upon the rate at which rental rates change, which is determined by the output elasticity of capital $\hat{\alpha}_j$, and the importance of new investment in accumulation $(1-\omega_j)$. This, in turn, implies that the size distribution of firms is not a Pareto distribution with coefficient 1 when $\hat{\alpha}_j > 0$. In this case the log rank log employee relationship is concave. In the following proposition we show that for $\hat{\alpha}_j$ smaller than one-half the larger $\hat{\alpha}_j$ the more reversion to the mean. That is the difference in average growth rates between small and large firms increases with $\hat{\alpha}_j$.[4]

Proposition 1 *Take two sequences* $\{A_t^1\}_{t=0}^{\infty}$ *and* $\{A_t^2\}_{t=0}^{\infty}$ *such that* $A_t^1 > A_t^2$. *Then,*

[4] In general, the growth rate of a firm can be expressed as a function of two state variables (productivity and capital). This implies that we cannot express growth rate solely as a function of firm size. Therefore, we relate both firm sizes and firm growth rates with ordered sequences of past shocks. In the data we identify these sequences as firm sizes.

for all t, the corresponding firm sizes satisfy, $\ln n_t^1 < \ln n_t^2$. *Furthermore,*

$$E_t\left(\ln n_{t+1}^1 - \ln n_t^1\right) \geq E_t\left(\ln n_{t+1}^2 - \ln n_t^2\right),$$

with equality when $\hat{\alpha} = 1$ *or* 0, *and* $\left|E_t\left(\ln n_{t+1}^1 - \ln n_t^1\right) - E_t\left(\ln n_{t+1}^2 - \ln n_t^2\right)\right|$ *increases with* $\hat{\alpha}$ *for* $\hat{\alpha} < \bar{\alpha}(\omega)$, *where* $\bar{\alpha}(\omega) > 1/2$.

Proof. Define

$$G(\hat{\alpha}, \omega, S) \equiv (1-\omega)(1-\hat{\alpha})(\omega + (1-\omega)\hat{\alpha})^{S-1},$$

and suppress the industry j subscript for convenience. Then the sum in (15) becomes

$$\sum_{S=1}^{\infty} (1-\omega)\hat{\alpha} G(\hat{\alpha}, \omega, S) \ln(A_{t-S}).$$

Past shocks affect current firm growth rates only through this term. The effect of a particular shock S periods away affects growth rates according to the value of $(1-\omega)\hat{\alpha}G(\hat{\alpha}, \omega, S)$. Hence, since this term is positive, firms in industries that received large shocks in period S, and so are small, grow faster than firms in industries that received low shocks, and so are large, in the same period. The larger this term the bigger the effect of the shock. Hence given a sequence of shocks (that determines the size of a firm), a firm that received large shocks on average grows faster, and more rapidly the larger $(1-\omega)\hat{\alpha}G(\hat{\alpha}, \omega, S)$ for all positive integers S. Therefore, to understand the effect of $\hat{\alpha}$ on growth rates we only need to calculate

$$\begin{aligned}
&\frac{\partial (1-\omega)\hat{\alpha}G(\hat{\alpha}, \omega, S)}{\partial \hat{\alpha}} \\
= \;& (1-\omega)^2 (\omega + (1-\omega)\hat{\alpha})^{S-2} \left[(1-\hat{\alpha})(\omega + (1-\omega)\hat{\alpha}) + \hat{\alpha}(S(1-\omega)(1-\hat{\alpha}) - 1)\right].
\end{aligned}$$

Notice that if $\hat{\alpha} = 0$, $(1-\omega)\hat{\alpha}G(\hat{\alpha}, \omega, S) = 0$ and so the growth rate of firms does not depend on past growth rates. Since $G(1, \omega, S) = 0$, the same is true for $\hat{\alpha} = 1$. For any value of $\hat{\alpha}$ such that $0 < \hat{\alpha} < 1$, $(1-\omega)\hat{\alpha}G(\hat{\alpha}, \omega, S) > 0$ which proves the first claim in the proposition.

Clearly

$$(1-\omega)^2 \hat{\alpha} (\omega + (1-\omega) \hat{\alpha})^{S-2} > 0,$$

so to prove the second claim we just need to verify the sign of

$$(1-\hat{\alpha}) (\omega + (1-\omega) \hat{\alpha}) + \hat{\alpha} (S (1-\omega) (1-\hat{\alpha}) - 1) .$$

We can rewrite the term to get

$$(1-\hat{\alpha}) (\omega + (1-\omega) \hat{\alpha}) - \hat{\alpha} (\omega + (1-\omega) \hat{\alpha}) + (S-1) (1-\omega) (1-\hat{\alpha}) \hat{\alpha}. \tag{16}$$

Let $S = 1$, then the term becomes

$$(1-\hat{\alpha}) (\omega + (1-\omega) \hat{\alpha}) - \hat{\alpha} (\omega + (1-\omega) \hat{\alpha})$$

which is greater than zero if $\hat{\alpha} < 1/2$ for all ω. For $S > 1$, notice that

$$(S-1) (1-\omega) (1-\hat{\alpha}) \hat{\alpha} > 0$$

and so (16) is positive for all S if $\hat{\alpha} < 1/2$. Define $\bar{\alpha} (\omega, S)$ as the capital share such that (16) is equal to zero given S and ω. Since (16) is continuous, positive for $\hat{\alpha} = 0$ and negative for $\hat{\alpha} = 1$, $\bar{\alpha} (\omega, S)$ exists. Let $\bar{\alpha} (\omega) = \inf_S [\bar{\alpha} (\omega, S)]$, then $\bar{\alpha} (\omega) > 1/2$ by the argument above. ■

The log-linearity of the model was shown above to imply that the employment allocation across industries was constant over time. Combined with the result of the above proposition, this has strong implications on exit rates: whenever firm sizes grow on average, there is exit. In a more general model in which the labor allocation varies in equilibrium this result continues to hold as long as the elasticity of substitution in consumption of each good is not too large. This is sufficient to guarantee that the labor allocation to the industry does not change by as much as firm sizes. Moreover, the above proposition implies that, for capital shares below one-half, the higher the capital share, the faster the exit rate decreases with firm size.

Corollary 2 *Take two sequences* $\{A_t^1\}_{t=0}^{\infty}$ *and* $\{A_t^2\}_{t=0}^{\infty}$ *such that* $A_t^1 > A_t^2$. *Then*

$$E_t\left(Exit\ Rate_{t+1}^1\right) \geq E_t\left(Exit\ Rate_{t+1}^2\right),$$

with equality when $\hat{\alpha} = 1$ *or* 0, *and* $\left|E_t\left(Exit\ Rate_{t+1}^1\right) - E_t\left(Exit\ Rate_{t+1}^2\right)\right|$ *increases with* $\hat{\alpha}$ *for* $\hat{\alpha} < \bar{\alpha}(\omega)$, *where* $\bar{\alpha}(\omega) > 1/2$.

These implications for the relationship between capital shares, firm growth rates and exit can be tested directly using longitudinal data. In combination with the assumption that the distribution of firm sizes has converged to its long-run distribution, we can also test this implication with data on the size distribution of firms. Rossi-Hansberg and Wright (2004) showed that the combination of scale independent growth for a finite number of industries, combined with this form of entry and exit, is sufficient to generate an invariant distribution that satisfied Zipf's law: the size distribution is Pareto with coefficient one. If mean reversion in growth rates increases, it can also be established that the invariant distribution appears concave in such a plot: relative to Zipf's Law, there is a relative absence of very small, and very large, firms. These claims are proven in the following two propositions.

Proposition 3 *(Zipf's Law) If* $\hat{\alpha}_j = 0$ *and productivity shocks are permanent, or in the limit as* $\hat{\alpha}_j \to 1$ *when productivity shocks are i.i.d., the size distribution of firms converges to a Pareto distribution with shape coefficient one.*

Proof. See Rossi-Hansberg and Wright (2004) Proposition 4. ■

For values of α between zero and one, we can also characterize the invariant distribution of firm sizes. We begin by establishing the existence of a unique invariant distribution. The proof of the following proposition requires compactness of the space of firm sizes and productivity levels. We obtain this by directly assuming that log productivity levels lie in the compact set $\left[\ln \underline{A}, \ln \overline{A}\right]$ for some $\underline{A}$ suitably small and $\overline{A}$ suitably large, and that firm sizes are measured relative to trend (or equivalently that population growth is zero). Specifically, we assume that

$$\ln A_{t+1j} = \max\left[\ln \underline{A}_j, \min\left[\ln \bar{A}_j, \rho_j \ln A_{tj} + \varepsilon_{tj}\right]\right],$$

where ε_t is a random i.i.d. shock with c.d.f. G. As long as $\left|\rho_j\right| < 1$, the bounds on TFP could be derived from bounds on the innovations ε_t.

Proposition 4 *For any $\hat{\alpha}_j \in (0,1)$, there exists a unique invariant distribution over firm and productivity levels in sector j.*

Proof. The proof is independent for each sector so we drop j from the notation. The size of a firm at time $t+1$ is given by

$$\ln n_{t+1} = C' - \frac{1}{\alpha+\beta} \ln A_{t+1} - \hat{\alpha} \ln K_{t+1},$$

where C' is a constant that includes employment in the industry that we know is a constant fraction of the population. For simplicity, assume that the population size is fixed (alternatively, we could work with variations from trend). A similar equation holds for $\ln n_t$ and we know that investment is a constant fraction of production in the industry. Hence we can obtain an expression for $\ln K_{t+1}$ as a function of $\ln A_t$ and $\ln n_t$ for $\hat{\alpha} > 0$. Substituting above, we have that firm size is given by

$$\ln n_{t+1} = C - \frac{1}{\alpha+\beta}\left[\ln A_{t+1} - \omega \ln A_t\right] + \left(1 - (1-\omega)(1-\hat{\alpha})\right) \ln n_t.$$

Define the following bounded set in $\mathbb{R}^2$

$$\bar{S} = \left[\frac{C - \frac{\ln \bar{A} - \omega \ln \underline{A}}{\alpha+\beta}}{(1-\omega)(1-\hat{\alpha})}, \frac{C - \frac{\ln \underline{A} - \omega \ln \bar{A}}{\alpha+\beta}}{(1-\omega)(1-\hat{\alpha})}\right] \times \left[\ln \underline{A}, \ln \bar{A}\right],$$

and define the two dimensional transition function $g : \bar{S} \times \mathbb{R} \rightarrow \bar{S}$ by

$$g\left(\left(\ln n, -\ln A\right), \varepsilon\right) = \begin{cases} \ln n' = C - \frac{1}{\alpha+\beta}\left[\ln A' - \omega \ln A\right] + \left(1 - \left(1 - \omega\right)\left(1 - \hat{\alpha}\right)\right) \ln n \\ -\ln A' = -\max\left[\ln \underline{A}, \min\left[\ln \bar{A}, \rho \ln A + \varepsilon\right]\right] \end{cases}.$$

Then, the probability of a transition from a point s to a set S is given by

$$Q\left(s, S\right) = \int_{-\infty}^{\infty} \mathbf{1}_{g(s,\varepsilon)\in S} dG\left(\varepsilon\right).$$

For any function $f : \bar{S} \rightarrow \mathbb{R}$ define the operator T by

$$\left(Tf\right)\left(s\right) = \int_{\bar{S}} f\left(s'\right) Q\left(s, ds'\right) = \int_{-\infty}^{\infty} f\left(g(s, \varepsilon)\right) dG\left(\varepsilon\right).$$

Define also the operator T^*, that maps the probability of being in a set S next period given the current distribution, say λ, as

$$\left(T^*\lambda\right)\left(S\right) = \int_{\bar{S}} Q\left(s, S\right) \lambda\left(ds\right).$$

Since the set $\bar{S}$ is compact, we are able to use Theorem 12.12 in Stokey, Lucas and Prescott (1989) to prove that there exists a unique invariant distribution, if we can show that the transition probability function Q satisfies the Feller property, is monotone, and satisfies the mixing condition.

To see that it satisfies the Feller Property, note that the function g is continuous in $\ln n$, $\ln A$, and ε, since the minimum and maximum preserve continuity. Since g is continuous and bounded, $f\left(g(\cdot)\right)$ is continuous and bounded and therefore so is Tf. Hence T maps the space of bounded continuous functions into itself, $T : C(\bar{S}) \rightarrow C(\bar{S})$. To see that it is monotone, we need to prove that if $f : \bar{S} \rightarrow \mathbb{R}$ is a non-decreasing function, then so is Tf. But this follows from the fact that the first coordinate of g is non-decreasing in both $\ln n$ and $\ln A$, and that the second coordinate is non-decreasing in $\ln A$. Hence $f\left(g(\cdot, \varepsilon)\right)$ is non-decreasing in s and therefore so is Tf.

Finally, to show that it satisfies the mixing condition, we need to show that there exists $c \in \bar{S}$ and $\eta > 0$ such that

$$Q\left(\left(\frac{C - \frac{\ln \bar{A} - \omega \ln \underline{A}}{\alpha+\beta}}{(1-\omega)(1-\hat{\alpha})}, \ln \underline{A}\right), \left[c, \left(\ln N, \ln \bar{A}\right)\right]\right) \geq \eta,$$

and

$$Q\left(\left(\ln N, \ln \bar{A}\right), \left[\left(\frac{C - \frac{\ln \bar{A} - \omega \ln \underline{A}}{\alpha+\beta}}{(1-\omega)(1-\hat{\alpha})}, \ln \underline{A}\right), c\right]\right) \geq \eta.$$

Let $c = (0, 0)$. Since

$$g\left(\left(-\frac{\ln \bar{A} - \omega \ln \underline{A}}{\alpha + \beta}, \ln \underline{A}\right), \varepsilon\right) = \begin{cases} \ln \bar{n}\left(\bar{A}, \underline{A}, \varepsilon\right) \\ -\max\left[\ln \underline{A}, \min\left[\ln \bar{A}, \rho \ln \underline{A} + \varepsilon\right]\right] \end{cases},$$

where $\ln \bar{n}$ is a finite function that is continuous and decreasing in ε, there exists an ε' such that for all $\varepsilon \leq \varepsilon'$, $\ln \bar{n}\left(\bar{A}, \underline{A}, \varepsilon\right) > 0$. Let $\eta' = 1 - G(\varepsilon')$. Similarly there exists an ε'' such that for all $\varepsilon \leq \varepsilon''$, $-\max\left[\ln \underline{A}, \min\left[\ln \bar{A}, \rho \ln \underline{A} + \varepsilon\right]\right] > 0$. Let $\eta'' = 1 - G(\varepsilon'')$. The same set of arguments define probabilities to move from the upper right corner of the set to the lower-left half. Call the minimum of these probabilities η.Then $c = (0, 0)$ and η guarantee that the mixing condition holds.

Theorem 12.12 in Stokey, Lucas and Prescott (1989) then guarantees that there exists a unique invariant distribution, and that the iterates of T^* weakly converge to that invariant distribution. ■

For capital shares strictly between zero and one, we are able to establish that the invariant distribution of firms sizes has thinner tails than the Pareto distribution with coefficient one. We make this notion precise in the following definition and proposition.

Definition 5 *Let λ and ψ be densities on $\left[\underline{b}, \overline{b}\right]$. The density function λ has* ***thinner tails*** *than ψ if there exists $\underline{x}$ and $\overline{x} \in \left[\underline{b}, \overline{b}\right]$ such that for all $0 \leq x \leq \underline{x}$, $\lambda(x) \leq \psi(x)$, for all $\underline{x} \leq x \leq \overline{x}$, $\lambda(x) \geq \psi(x)$, and for all $\overline{x} \leq x \leq 1$, $\lambda(x) \leq \psi(x)$.*

Proposition 6 *If $\hat{\alpha} \in (0,1)$ the invariant distribution of firm sizes has thinner tails than the Pareto distribution with coefficient one.*

Proof. Denote the unique invariant probability measure of firm sizes (see Proposition 4) and TFP levels by $\lambda : \bar{\mathcal{S}} \to [0,1]$, where $\bar{\mathcal{S}}$ denotes the $\sigma-$algebra associated with $\bar{S}$. Then the distribution of firm sizes is given by

$$\lambda^n(\ln n) = \int_{\ln \underline{A}}^{\ln \bar{A}} \lambda\left(\ln n, \ln A\right) d\ln A,$$

and the distribution of TFP by

$$\lambda^A(\ln A) = \int_{\frac{C-\left(\ln \underline{A}-\omega \ln \bar{A}\right)(\alpha+\beta)}{(1-\omega)(1-\hat{\alpha})}}^{\frac{C-\left(\ln \bar{A}-\omega \ln \underline{A}\right)(\alpha+\beta)}{(1-\omega)(1-\hat{\alpha})}} \lambda\left(\ln n, \ln A\right) d\ln n.$$

Given that the stochastic process of TFP levels is independent of firm size we know that

$$\lambda\left(\ln n, \ln A\right) = \lambda^A(\ln A)\hat{\lambda}^n(\ln n, \ln A)$$

where $\hat{\lambda}^n(\ln n, \ln A)$ is the distribution of firm sizes given a TFP level $\ln A$. Since λ^n is an invariant distribution

$$\begin{aligned}
\lambda^n\left(\ln n\right) &= \int_{\ln \underline{A}}^{\ln \bar{A}} \left(T^* \lambda^A \lambda^n\right)\left(\ln n\right) d\lambda^A\left(\ln A\right) \\
&= \int_{-\infty}^{\infty} \int_{\ln \underline{A}}^{\ln \bar{A}} \lambda^n\left(gr(\ln n, \ln A, \hat{\alpha}) - \ln n\right) d\lambda^n\left(\ln A\right) dG\left(\varepsilon\right),
\end{aligned}$$

where $gr(n, A, \hat{\alpha})$ denotes the log firm size growth rate. We know that $gr(n, A, \hat{\alpha}) = C' - \frac{1}{\beta}\left[\ln A' - \ln A\right]$ for $\hat{\alpha} = 0$, and

$$gr(n, A, \varepsilon; \hat{\alpha}) = C - \frac{1}{\alpha+\beta}\left[\ln A' - \omega \ln A\right] + \left(1 - \left(1-\omega\right)\left(1-\hat{\alpha}\right)\right)\ln n$$

for $\hat{\alpha} \in (0,1)$. Proposition 1 tells us that

$$\frac{dgr(n, A, \varepsilon; \hat{\alpha})}{dn} < 0.$$

Proposition 3 shows that the Pareto distribution is the invariant distribution of firm sizes for $\hat{\alpha} = 0$, call the corresponding $\ln n$ distribution λ^P. Then, for n small enough, we know that

$$\begin{aligned}
\lambda^P(\ln n) &= \int_{\ln \underline{A}}^{\ln \bar{A}} \left(T^{*0}\lambda^A\lambda^P\right)(\ln n)\, d\lambda^A(\ln A) \\
&= \int_{-\infty}^{\infty}\int_{\ln \underline{A}}^{\ln \bar{A}} \lambda^P\left(gr(\ln n, \ln A, \varepsilon; 0) - \ln n\right) d\lambda^A(\ln A)\, dG(\varepsilon) \\
&> \int_{-\infty}^{\infty}\int_{\ln \underline{A}}^{\ln \bar{A}} \lambda^P\left(gr(\ln n, \ln A, \varepsilon; \hat{\alpha}) - \ln n\right) d\lambda^A(\ln A)\, dG(\varepsilon),
\end{aligned}$$

Hence the Pareto distribution is not the invariant distribution for $\hat{\alpha} > 0$, and the operator T^* maps the Pareto distribution in distributions with lower tails (the case for intermediate and high $\ln n$ are analogous). ■

In Proposition 1, we found that the degree of mean reversion in firm growth rates was increasing in the capital share for capital shares less than one half. Firm size distributions should then display thinner tails the greater is the capital share within this range. The intuition for this result is that the greater the degree of mean reversion in the growth process, the more the mass of firms is shifted towards the center of the distribution. This intuition is, however, difficult to formalize, but can be verified numerically. As capital shares rise above one half, the same logic implies that firm size distributions become less concave, as mean reversion in firm growth rates diminishes. This links our results on growth rates into result on the shape of the size distribution of firms.

ROBUSTNESS OF THE MECHANISM

In the introduction we argued that it is essential that any proposed explanation for the documented patterns in firm dynamics and size distribution be robust to the wide variety of differences in institutions and market structures for which these patterns have been observed. In this section, we establish that the mechanism described above in a particular setup survives generalization to environments in which the specification of firm costs are different (so that firms respond in the opposite way to productivity shocks), to the addition of human capital, to the introduction of firm level heterogeneity, to environments in which governments levy taxes of various kinds, and to an environment in which competition amongst firms is monopolistic. In each case, we show how the general pattern of mean reversion in industry specific factor stocks leads to mean reversion in firms sizes.

Firm Costs

The basic mechanism of our paper relies on mean reversion in the stock of industry specific factors of production, which is the natural result of concavity of the production function. It is this mean reversion in turn that leads to the mean reverting characteristics that we emphasized for firm dynamics and size distributions.

Nothing about this argument depends upon the qualitative relationship between the relative stock of factors, and the relative size of the firm. In the model presented above, we assumed for simplicity that the firms cost structure combined decreasing returns to scale with a fixed cost denominated in terms of the firm's output. This combination implied that the output of the firm was constant, so that firms reduced employment (and hence size in terms of employment) when productivity increased, or when the stock of specific factors grew. In other words, reversion to the mean in the stock of specific factors *from above*, produced reversion to the mean in firm sizes *from below*.

Changes to the exact specification of a firms cost structure can allow output at a firm to vary and may change the qualitative relationship between factor supplies and firm sizes, without changing the implication that mean reversion in factor supplies produces mean reversion in firm sizes. For example, suppose that fixed costs took the form of an overhead cost specified as an amount ϕ_j of the weighted average of inputs specified by the production function, so that

$$y_{tj} = A_{tj} \left(k_{tj}^{\alpha_j} n_{tj}^{\beta_j} - \phi_j \right).$$

In this case, the weighted average input mix of the firm would be constant and independent of productivity shocks, firm output would increase with productivity shocks, and firm employment would be given by

$$n_{tj} = \frac{N_{tj}}{\mu_{tj}} = \left[\frac{\phi_j}{(1 - \alpha_j - \beta_j)} \right]^{\frac{1}{\alpha_j + \beta_j}} K_{tj}^{-\frac{\alpha_j}{\alpha_j + \beta_j}} N_{tj}^{\frac{\alpha_j}{\alpha_j + \beta_j}}.$$

If the stock of industry specific capital increases, the firm responds by using more of this capital and decreasing it's employment level, exactly as before.

Changes in the specification of the cost structure also have the potential to reverse the qualitative relationship between factor supplies and firm size. To see this, assume as before that each firm in industry j at time t produces output according to

$$y_{tj} = A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j},$$

where $\alpha_j + \beta_j < 1$ so that firms face a technology with decreasing returns to scale. Now, however, assume that hiring n_{tj} workers entails an additional managerial cost of $F_j n_{tj}^{\gamma}$, so that the problem of the firm is to maximize profits

$$\max_{k_{tj}, n_{tj}} \Pi \equiv \max_{k_{tj}, n_{tj}} A_{tj} k_{tj}^{\alpha_j} n_{tj}^{\beta_j} - r_{tj} k_{tj} - w_{tj} n_{tj} - F_j n_{tj}^{\gamma}.$$

Clearly, if $\gamma_j = 0$, we have the same case studied above; $\gamma_j < 1$ by assumption. Taking first order conditions and allowing for free entry and exit so that profits are zero implies

$$(1 - \alpha_j - \beta_j)\, y_{tj} = (1 - \gamma_j)\, F_j n_{tj}^{\gamma_j}.$$

Since all firms producing in industry j are identical, equilibrium in factor markets implies

$$\left(1-\alpha_j-\beta_j\right) A_{tj}\left(\frac{K_{tj}}{\mu_{tj}}\right)^{\alpha_j}\left(\frac{N_{tj}}{\mu_{tj}}\right)^{\beta_j}=\left(1-\gamma_j\right) F_j\left(\frac{N_{tj}}{\mu_{tj}}\right)^{\gamma_j},$$

which gives the following expressions for the number of firms in the industry

$$\mu_{tj}=\left[\frac{A_{tj}\left(1-\alpha_j-\beta_j\right)}{\left(1-\gamma_j\right) F_j}\right]^{\frac{1}{\alpha_j+\beta_j-\gamma_j}} K_{tj}^{\frac{\alpha_j}{\alpha_j+\beta_j-\gamma_j}} N_{tj}^{\frac{\beta_j-\gamma_j}{\alpha_j+\beta_j-\gamma_j}\,j},$$

and the size of a the typical firm in the industry

$$n_{tj}=\frac{N_{tj}}{\mu_{tj}}=\left[\frac{\left(1-\gamma_j\right) F_j}{A_{tj}\left(1-\alpha_j-\beta_j\right)}\right]^{\frac{1}{\alpha_j+\beta_j-\gamma_j}} K_{tj}^{-\frac{\alpha_j}{\alpha_j+\beta_j-\gamma_j}} N_{tj}^{\frac{\alpha_j}{\alpha_j+\beta_j-\gamma_j}}.$$

These equations are obviously analogous to the case considered above with a pure fixed cost. The main differences are that now both employment and output respond to changes in productivity levels and factor supplies. Moreover, the direction of the change can differ: for $\gamma_j < \alpha_j+\beta_j$, the behavior of employment is as before, declining with increases in productivity levels; for $\gamma_j > \alpha_j + \beta_j$ this pattern is reversed and higher productivity leads to larger firm sizes. In either case, the main properties for firm growth and exit rates, and the size distribution, are preserved: regardless of whether firms in industries with large capital stocks are large or small, for intermediate capital shares they revert to the mean, with the rate of mean reversion increasing with the curvature of the production function; for industries in which the capital share is zero, firm growth rates are characterized by Gibrat's Law, and firm size distributions converge to Zipf's Law.

Within Industry Firm Heterogeneity

In the theory presented above, we abstracted from heterogeneity amongst firms *within* an industry in order to focus our attention on heterogeneity *across* industries. This allowed us to emphasize the contribution of the accumulation of industry specific capital to the evolution of firm sizes. Some authors have argued that there exist differences in firm sizes even within narrowly defined industries. While this may be caused by aggregation (data is rarely available beyond the three or four digit SIC levels), it is probable that some firm specific heterogeneity remains. In this section we demonstrate how firm specific heterogeneity can be added to our framework, and show that it does not change the key empirical implications of our theory for the differences in firm dynamics and size distributions across industries.

Consider the model of Section 2. Suppose that after having decided to produce in a period (that is, after paying the fixed cost F) each firm $i \in [0, \mu_{jt}]$ in industry j at time t then observes a firm specific productivity shock θ_i. This shock is assumed to be iid over time and firms and industries within a sector. After observing this shock, the firm can then hire labor n_i and industry-j-specific capital k_i to produce output according to

$$y_i = A\theta_i k_i^{\alpha} n_i^{\beta}, \tag{17}$$

where from now on we suppress time and industry subscripts.

To see how this affects the results, we consider once again the social planners problem. To begin, suppose that the planner has decided that there are μ firms in the industry employing N workers. The amount of industry specific capital is fixed at K. The planner then observes the identities of the firms that receive each productivity shock. The problem of the planner is then to allocate factors across firms in the industry to maximize industry output

$$\int_0^{\mu} A\theta_i k_i^{\alpha} n_i^{\beta} di,$$

subject to

$$\int_0^{\mu} k_i di \leq K, \qquad \int_0^{\mu} n_i di \leq N.$$

We assume that we can index the productivity shock by the unit interval with density ϕ and that the appropriate LLN holds for continua of iid random variables. Then this problem becomes one of maximizing

$$A\mu \int_0^1 \theta_i k_i^\alpha n_i^\beta \phi(di),$$

subject to

$$\mu \int_0^1 k_i \phi(di) \leq K, \qquad \mu \int_0^1 n_i \phi(di) \leq N.$$

The first order conditions for this problem imply a relative allocation of factors of

$$\frac{k_i}{k_j} = \left(\frac{\theta_i}{\theta_j}\right)^{\frac{1}{1-\alpha}}, \qquad \frac{n_i}{n_j} = \left(\frac{\theta_i}{\theta_j}\right)^{\frac{1}{1-\beta}},$$

and relative outputs

$$\frac{y_i}{y_j} \equiv \frac{A\theta_i k_i^\alpha n_i^\beta}{A\theta_j k_j^\alpha n_j^\beta} = \left(\frac{\theta_i}{\theta_j}\right)^{\frac{1-\alpha\beta}{(1-\alpha)(1-\beta)}}.$$

That is, firms within an industry with a higher shock use more of both inputs and produce more output. Actual amounts used in each firm can be determined from the resource constraint so that

$$k_i = \frac{\theta_i^{\frac{1}{1-\alpha}}}{\mu \int_0^1 \theta_i^{\frac{1}{1-\alpha}} \phi(di)} K, \qquad n_i = \frac{\theta_i^{\frac{1}{1-\beta}}}{\mu \int_0^1 \theta_i^{\frac{1}{1-\beta}} \phi(di)} N.$$

With these results, we can characterize the level of output in the industry given the initial choice of the number of firms μ, the choice of labor N, the aggregate productivity shock, and previously accumulated capital K as

$$\begin{aligned} A\mu \int_0^1 \theta_i k_i^\alpha n_i^\beta \phi(di) &= A \frac{\left(\int_0^1 \theta_i \theta_i^{\frac{\alpha}{1-\alpha}} \theta_i^{\frac{\beta}{1-\beta}} \phi(di)\right)}{\int_0^1 \theta_i^{\frac{1}{1-\alpha}} \phi(di) \int_0^1 \theta_i^{\frac{1}{1-\beta}} \phi(di)} K^\alpha N^\beta \mu^{1-\alpha-\beta} \\ &\equiv \hat{A} K^\alpha N^\beta \mu^{1-\alpha-\beta}. \end{aligned}$$

From this equation, it is easy to see that the form of the industry production function is exactly the same as for the original problem, and consequently that the choices of N and μ are analogously determined.

Clearly, the addition of an iid productivity shock has no effect on the mean growth and exit rates of firms in that industry. Consequently, the model has the same implications for growth and exit at the sector level. Further, the distribution of *average* firm sizes is unchanged, and so the relationship between factor intensities and the shapes of the firm size distribution is unchanged. One implication that can be affected is the range of cases under which Zipf's Law exactly holds: when the conditions of Proposition 3 hold, we observe Zipf's Law for average firm sizes, but only for actual firm sizes if either all firms are identical within an industry, or if the distribution within an industry is also Pareto with coefficient one. We might think of the latter as being produced by a similar mechanism as the one laid out in this paper, working through firm specific capital.

Industry Specific Human Capital

Our mechanism emphasizes the accumulation of industry specific factors which we have, for simplicity, referred to as capital. When we take the implications of our model to the data we also focus on capital as an industry specific factor. In some industries, there is also a great deal of industry specific *human* capital. In this subsection we show that, as long as industry specific human capital is produced primarily using itself as an input, the implications of our model for industry specific *physical* capital are unchanged. This underscores the point that it is the role of non-linear accumulation of industry specific factors that drives our mechanism.

To see this, suppose that in each period t the household begins with stocks of industry-j-specific human capitals H_{tj}, a fraction u_{tj} of which can be devoted to production with the rest reserved for accumulating more human capital according to

$$H_{t+1j} = \left[1 + B\left(1 - u_{tj}\right)\right] H_{tj}.$$

The fraction of human capital devoted to work $u_{tj}H_{tj}$ has an output elasticity of γ_j

so that the resource constraint for an industry is now

$$C_{tj} + X_{tj} + F_j \mu_{tj} \leq A_{tj} K_{tj}^{\alpha_j} N_{tj}^{\beta_j} \left(u_{tj} H_{tj}\right)^{\gamma_j} \mu_{tj}^{1-\alpha_j-\beta_j-\gamma_j}.$$

It is easily verified that the optimal choice of the number of firms now implies

$$\mu_{tj} = \left[\frac{A_{tj}\left(1-\alpha_j-\beta_j-\gamma_j\right)}{F_j}\right]^{\frac{1}{\alpha_j+\beta_j+\gamma_j}} K_{tj}^{\frac{\alpha_j}{\alpha_j+\beta_j+\gamma_j}} N_{tj}^{\frac{\beta_j}{\alpha_j+\beta_j+\gamma_j}} \left(u_{tj} H_{tj}\right)^{\frac{\gamma_j}{\alpha_j+\beta_j+\gamma_j}},$$

while the optimal allocation of human capital between work and study is constant $u_{tj} = u_j^*$. But given the linearity of the production function for human capital, this implies that labor input continues to grow at a constant rate, so that the evolution of the number of firms and their size is driven solely by the direct effect of productivity shocks and their indirect effect through the accumulation of physical capital. Once again it is the share of physical capital in industry value added

$$\hat{\alpha}_j = \frac{\alpha_j}{\alpha_j + \beta_j + \gamma_j},$$

that drives the evolution process of firms sizes.

Institutions, Government and Taxes

An important aspect of the results that we cited in the introduction is that they have been found to hold across a wide variety of time periods and countries. One question that faces any potential mechanism consistent with these facts is whether or not the explanation is robust to the variation in government policies that is observed over time and across countries. In this section, we demonstrate that our mechanism is robust to several forms of government taxation, for which there is variation over countries. To the extent that differences in institutions manifest themselves as tax wedges, this is also sufficient to establish that the mechanism is robust to some differences in institutions as well.

One form of government intervention that is obviously consistent with our mechanism is the presence of taxes and regulations that are purely wasteful and which have a fixed effect on a firm. If we let the size of such an intervention be given by T_j which plausibly can vary by industry, then analogous reasons to those presented above imply that the size of the firm is given by

$$n_{tj} = \frac{N_{tj}}{\mu_{tj}} = \left[\frac{F_j + T_j}{A_{tj} \left(1 - \alpha_j - \beta_j\right)} \right]^{\frac{1}{\alpha_j + \beta_j}} K_{tj}^{-\hat{\alpha}_j} N_{tj}^{\hat{\alpha}_j}.$$

As the effect of this intervention is simply to increase fixed costs, it acts in the same way as fixed costs. Moreover, as our mechanism was independent of the size of fixed costs, none of the implications of the model are changed.

Taxes on factor payments are similarly straightforward to incorporate. As labor is supplied inelastically, taxes on wage payments have no effect on labor supply. As a result we focus on taxes on capital income. It is a well known result of log-linear models of this sort that capital income taxes in such a model have an analogous effect to a reduction in the discount rate of the households in the economy. If we allow capital income taxes to vary by industry, τ_j, then it is as though the household discounts consumption of each good differently over time by the amount $\delta_j = \delta \left(1 - \tau_j\right)$.

Changes in the discount rate do not affect, for a given supply of factors, the size of the firm or the number of firms in the industry. Changes in capital taxes that act as though they are changing the discount rate affects the level of the supply of factors to the industry.

Monopolistic competition

The previous model uses an extremely simple theory of the firm to derive conclusions on the size distribution of firms. In this section we use a different theory of the firm to show that the conclusions derived above are not specific to that particular theory of the organization of production in firms. For this we use the Dixit-Stiglitz monopolistic competition model with taste for variety. In this model substitution for varieties in the same industry limits demand for a particular variety in an industry and therefore determines the size of the firm. The model includes naturally the two margins we have emphasized so far, the number of firms in an industry and the size of these firms. We need to present a version of this theory where both margins react to factor accumulation and productivity shocks. In particular, we need a version of the theory that includes the two factors that we introduced in the model above, labor and physical capital. As above, labor is mobile across all sectors. Now, capital is specific to an industry but mobile across varieties within that industry.

Households.—

As above, we assume that there are J industries divided into sectors with similar technologies. Now, however, we assume that each industry consists of a continuum of potential varieties which we index by ϖ. Households provide labor and industry-specific (but *not* variety-specific) capital to each variety within an industry. Output of each variety D_{tj}^{ϖ} is combined by the household using a constant elasticity of substitution production function with parameter σ_j to produce an aggregate good that is used for both final consumption and investment.

That is, the problem of a consumer is to purchase goods and accumulate industry specific capitals to maximize lifetime utility, or

$$\max_{D_{tj}^{\varpi}, N_{tj}, C_{tj}, X_{tj}} (1-\delta) E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \left(\sum_{j=1}^{J} \theta_j \ln \left(\frac{C_{tj}}{N_t} \right) \right) \right]$$

subject to

$$E_0 \left[\sum_{t=0}^{\infty} \sum_{j=1}^{J} \int_{0 \leq \varpi \leq \Omega_{tj}} p_{tj}^{\varpi} D_{tj}^{\varpi} d\varpi \right] \leq E_0 \left[\sum_{t=0}^{\infty} \sum_{j=1}^{J} p_{tj}^{\varpi} \left(w_{tj} N_{tj} + r_{tj} K_{tj} \right) \right],$$

$$K_{t+1j} = K_{tj}^{\omega_j} X_{tj}^{1-\omega_j},$$

for all t and all j,

$$C_{tj} + X_{tj} \equiv E_{tj} \leq \left\{ \int_{0 \leq \varpi \leq \Omega_{tj}} \left(D_{tj}^{\varpi}\right)^{\frac{\sigma_i - 1}{\sigma_i}} d\varpi \right\}^{\frac{\sigma_i}{\sigma_i - 1}},$$

for all t and all j, where E_{tj} is total demand for the final good from industry j, and

$$\sum_{j=1}^{J} N_{tj} \leq N_t,$$

all t. The consumer takes as given the prices of intermediate inputs and factors, as well as the range of varieties of goods available.

In order to solve the firms problem below, it is useful to record that the first order conditions of the consumers problem with respect to a variety implies a demand for variety ϖ in industry j of

$$D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right) = E_{tj}^{\varpi} \frac{\left(p_{tj}^{\varpi}\right)^{-\sigma_j}}{\int\limits_{0 \leq \varpi \leq \Omega_{tj}} \left(p_{tj}^{\varpi}\right)^{1-\sigma_j} d\varpi},$$

where Ω_{tj} is the measure of varieties that make positive profits and therefore produce in equilibrium in industry j at time t, which consumers take as given.

Firms and industry equilibrium.—

Firms producing variety ϖ use a constant returns to scale Cobb-Douglas technology with labor and physical capital as factors of production, given by

$$Y_{tj}^{\varpi} = A_{tj} \left(K_{tj}^{\varpi}\right)^{\alpha_j} \left(N_{tj}^{\varpi}\right)^{1-\alpha_j}$$

The first stage of the problem of the firm is to minimize costs,

$$\begin{aligned} C\left(r_{tj}, w_{tj}, D_{tj}^{\varpi}, F_j\right) &\equiv \min_{K_{tj}^{\varpi}, L_{tj}^{\varpi}} r_{tj} K_{tj}^{\varpi} + w_{tj} N_{tj}^{\varpi} \\ s.t.\ D_{tj}^{\varpi} + F_j &= A_{tj} \left(K_{tj}^{\varpi}\right)^{\alpha_j} \left(N_{tj}^{\varpi}\right)^{1-\alpha_j}, \end{aligned}$$

where D_{tj}^{ϖ} is the quantity consumed and invested of the good and F_j is a fixed cost of production in variety ϖ goods. The first order conditions of this problem are given by

$$r_{tj} = \lambda_{tj} \frac{\alpha_j \left(D_{tj}^{\varpi} + F_j\right)}{K_{tj}^{\varpi}}, \qquad w_{tj} = \lambda_{tj} \frac{(1-\alpha_j)\left(D_{tj}^{\varpi} + F_j\right)}{N_{tj}^{\varpi}},$$

where λ_{tj} is the multiplier from the firms problem. Combining these conditions with the technology constraint yields

$$\lambda_{tj} = \frac{1}{A_{tj}} \left(\frac{r_{tj}}{\alpha_j}\right)^{\alpha_j} \left(\frac{w_{tj}}{1-\alpha_j}\right)^{1-\alpha_j},$$

and so

$$C\left(r_{tj}, w_{tj}, D_{tj}^{\varpi}, F_j\right) = \frac{D_{tj}^{\varpi} + F_j}{A_{tj}} \left(\frac{r_{tj}}{\alpha_j}\right)^{\alpha_j} \left(\frac{w_{tj}}{1-\alpha_j}\right)^{1-\alpha_j}.$$

Notice that average costs $C\left(r_{tj}, w_{tj}, D_{tj}^{\varpi}, F_j\right)/D_{tj}^{\varpi}$ are a decreasing function of D_{tj}^{ϖ}.

The second stage of the firm problem is to maximize profits

$$\prod\left(r_{tj}, w_{tj}, F_j\right) = \max_{p_{tj}^{\varpi}} D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right) p_{tj}^{\varpi} - C\left(r_{tj}, w_{tj}, D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right), F_j\right),$$

where $D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right)$ is derived from the consumers problem.

The first order conditions of the firm problem then imply that

$$p_{tj}^{\varpi} = \lambda_{tj} \frac{\sigma_j}{\sigma_j - 1}.$$

Hence in equilibrium the quantity produced by firms is given by

$$D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right) = \frac{E_{tj}}{\Omega_{tj}\lambda_{tj}} \frac{\sigma_j - 1}{\sigma_j} = \frac{\sigma_j - 1}{\sigma_j} \frac{E_{tj} A_{tj}}{\Omega_{tj}} \left(\frac{\alpha_j}{r_{tj}}\right)^{\alpha_j} \left(\frac{1-\alpha_j}{w_{tj}}\right)^{1-\alpha_j}$$

and

$$\prod\left(r_{tj}, w_{tj}, F\right) = \frac{E_{tj}}{\sigma_j \Omega_{tj}} - F_j \lambda_{tj}.$$

Zero profits then implies that the number of varieties (or firms since only one firm produces each variety) is given by

$$\Omega_{tj} = \frac{E_{tj}}{\sigma_j F_j \lambda_{tj}},$$

and so

$$D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right)=F_j\left(\sigma_j-1\right).$$

The equilibrium conditions in factor markets are given by

$$K_{tj}=\Omega_{tj}\left(\lambda_{tj}\frac{\alpha_j\left(D_{tj}^{\omega}+F_j\right)}{r_{tj}}\right)=E_{tj}\frac{\alpha_j}{r_{tj}},\qquad N_{tj}=E_{tj}\frac{(1-\alpha_j)}{w_{tj}},$$

which implies that

$$\lambda_{tj}=\frac{E_{tj}}{A_{tj}}K_{tj}^{-\alpha_j}N_{tj}^{\alpha_j-1}.$$

Hence

$$\Omega_{tj}=\frac{E_{tj}}{\sigma_j F\lambda_{tj}}=\frac{A_{tj}}{\sigma_j F_j}K_{tj}^{\alpha_j}N_{tj}^{1-\alpha_j}.$$

Output in the industry is given by

$$Y_{tj}=\Omega_{tj}D_{tj}^{\varpi}\left(p_{tj}^{\varpi}\right)=\frac{\sigma_j-1}{\sigma_j}A_{tj}K_{tj}^{\alpha_j}N_{tj}^{1-\alpha_j}.$$

Notice that this function is constant returns to scale, with TFP given by

$$TFP_{tj}=\frac{\sigma_j-1}{\sigma_j}A_{tj}.$$

The size of firms in terms of employees is given by

$$N_{tj}^{\omega}=\frac{F_j\sigma_j}{A_{tj}}K_{tj}^{-\alpha_j}N_{tj}^{\alpha_j},$$

which has a very similar form to the one derived for the case of perfect competition above. As a result, the model has identical implications for the dynamics and size distribution of firm sizes.

Capital accumulation, labor allocation and firm sizes.—

All that remains is to calculate the accumulation decisions of agents. Although this can be done directly from the agents decision problem, it is instructive to compute them in an analogous way to the allocations for the perfectly competitive economy discussed above. Although the welfare theorems do not hold for this economy, the fact that the markup of these monopolistic firms is constant combined with the log-linearity of the model means that the equilibrium allocations can be obtained as the solution of an *equivalent optimum problem* that is identical to the social planners problem, used above, except that the resource constraint is now

$$C_{tj} + X_{tj} \leq \frac{\sigma_j - 1}{\sigma_j} A_{tj} K_{tj}^{\alpha_j} N_{tj}^{1-\alpha_j} \equiv Y_{tj},$$

for all t and j (see Chapter 18 of Stokey, Lucas and Prescott (1989) for another example of this pseudo-economy approach). As before, the solution of this model has the household accumulating a fixed proportion of the output of each industry as industry specific capital

$$X_{tj} = sY_{tj}.$$

The allocation of labor to work in each industry is fixed at the same levels as before.

From these results it is straightforward to show that the evolution of firm sizes in the model with monopolistic competition is identical (given that in this case $\alpha+\beta = 1$) to the evolution of firm sizes in the model with perfect competition. In particular, analogues of Propositions 1, 3, 4, and 6 and of Corollary 2 continue to hold.

EVIDENCE ON SCALE DEPENDENCE BY SECTOR

The model above has several empirical implications that are consistent with findings in the empirical literature. Firm growth and exit rates decline with size, and the size distribution has thinner tails than the Pareto with shape coefficient one. On top of this, in our theory the degree of reversion to the mean in capital stocks, and therefore in firms sizes, is determined by the degree of diminishing returns in capital. A very low capital share, or a capital share close to one, imply a low degree of diminishing returns in capital and, therefore, a low degree of reversion to the mean in firm sizes. As the capital share increases from zero, or decreases from one, the degree of diminishing returns in capital increases as does the reversion to the mean in firm sizes. Proposition 1 shows that the degree of mean reversion in firm sizes reaches a maximum for some capital share greater than one half. This implication of the model implies that the degree of mean reversion in growth rates, the degree of scale dependence in exit rates, and the thinness of the tails of the size distribution, are intrinsically determined by the importance of industry specific factors in technology. In this section we contrast this implication with the data.

Data

We have investigate the variation in scale dependence across sectors using data on growth rates and the distribution of firm sizes. We purchased two data sets from the US Census Bureau. The first is a dataset from the Statistics of US Businesses (SUSB) program on establishment size distributions by sector at the two digit SIC level for 1990 and three digit NAICS level for 2000. These data are constructed from a number of sources including the annual County Business Profile (CBP) data files. The second dataset, from the Business Information Tracking System (BITS), contains data on growth rates of establishments between 1990 and 2000, and deaths of establishments by size category for 1995-1996. These new data sets have several advantages for our purposes in comparison with the publicly available data sources. First, they provide the number of firms per size category for the finest size categories that the US Census will release given the confidentiality restrictions. Because of our emphasis on the shape of the size distribution, this level of detail is crucial. Previous analysis of the size distribution of firms have, to our knowledge, used data for much larger size bins or only for a couple of sectors. Second, it includes all sectors in the private non-farm US economy, including both manufacturing and services. This is important for our study given that we want to understand the effect of sectoral differences in capital shares on the size distribution of firms. Variations in capital shares are much larger across service and manufacturing sectors than within them. Third, the data refers to establishment sizes, and not enterprise sizes, which as we argued before makes a difference for large enterprises with several plants. The unique aspect of the longitudinal dataset is that it tracks the size of firms for several years, and, for exiting firms, for three years before they exit.

We also need to calculate capital shares. We do this using the Bureau of Economic Analysis (BEA) Industry Accounts. We use data on labor costs and value added at basic prices to construct labor shares. We then construct capital shares as one minus the labor share. This implies that the capital shares we use include everything that is not classified as labor. Human capital, even though it can be sector specific, is *not* accounted for in the capital share. This, however, does not create a problem if human capital is accumulated approximately linearly, as we argued in the previous section. The reason is that in this case linear accumulation implies that the rate at which human capital is accumulated is constant even with a technology that exhibits diminishing returns in human capital. There are two potential problems with the capital shares we compute. First, the capital shares include land shares. Land is not an industry specific factor, but as its share is usually small, this should have a negligible effect on the capital shares we use. Second, we are using the capital share in value added, but our theory is abstracting from the use of intermediate inputs. To address the latter concern, we also present results with capital shares adjusted for the share of value added and the share of materials purchased from the same industry.

Growth Rates

We begin by examining the growth rates of surviving firms. As a first step, consider an example with two sectors. Educational services is a very labor intensive sector with a capital share of 0.054, while manufacturing is much more capital intensive with a capital share of 0.397. If the theory is consistent with the data, given that manufacturing is more capital intensive, we should see growth rates of manufacturing firms decline faster with size than growth rates of firms in the educational sector.

Figure 5: Firm Growth Rates by Industry, 1990-2000

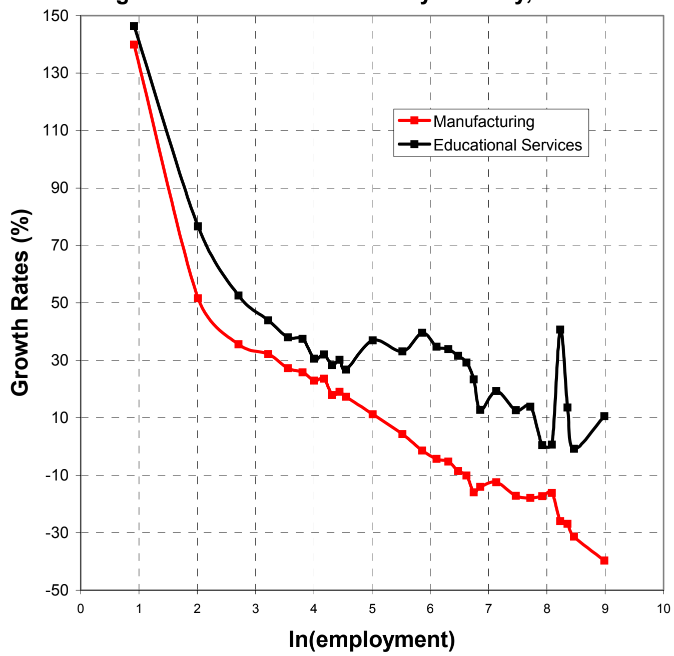


Figure 5 illustrates that this is the case, and shows that the differences are very large over a period of ten years. Not only do small firms grow faster than large firms in both sectors, but the scale dependence is significant for the entire range of firm sizes. The difference between the growth rates in these two sectors increases with firm size and is, for the largest firms, more than 40 per-cent.

This evidence is not particular to the pair of sectors in the example. Most sectors have capital shares smaller than one half. This implies that the most powerful prediction of the theory is that scale dependence in growth rates decreases with capital shares for sectors with capital shares smaller than one half. We examine next the implication of our theory that scale dependence in growth rates increases with capital shares (denoted by α_j) for all industries with $\alpha_j < 1/2$. We use data on the growth of firms, g_j, in a particular size category, x_j, and estimate the following regression:

$$\ln(1+g_j) = \tilde{a}_j + \tilde{b}\ln x_j + \hat{e}\alpha_j \ln x_j + \tilde{\varepsilon}_{tj}.$$

This amounts to fitting an exponential trend where the parameter varies linearly with capital shares by sector. We estimate this relationship using weighted least squares to take into account the fact that there are many more firms in the smaller size categories. We calculate the weights using data on the number of firms in each size category. The theory predicts that the estimate of $\tilde{e}$ should be negative and significant. The estimate of $\tilde{e}$ is presented in the first column of Table 1. The third column of Table 1 presents the result of a similar exercise fitting a power function instead of an exponential. Given the largest firm size in our sample, a larger (in absolute value) coefficient implies more scale dependence for all firm sizes. The results in Table 1 show that scale dependence increases significantly with sectoral capital shares.

Table 1

	Exponential 1990-2000	Exponential 1990-2000 (adjusted)	Power 1990-2000	Power 1990-2000 (adjusted)
$\tilde{e}$	-0.0965	-0.1303	-0.2638	-0.3503
Standard error	0.0273	0.0345	0.0195	0.0250
P-value	0.0004	0.0002	3.42×10^{-38}	3.46×10^{-41}

As mentioned before, the capital shares have been calculated as 1 minus the share of labor compensation in value added. Given that materials are an important fraction of gross output in an industry, this may result in capital shares that are too large relative to the ones in gross output. Since our theory does not include materials, the capital share in the theory refers to the capital share in gross output. To address these concerns we calculated the share of value added plus the share of inputs originating from the same sector using the input-output data provided by the BEA. We then multiply this share by the capital share to obtain an adjusted capital share. If all intermediate inputs originated in the same sector, the original capital shares would equal the adjusted capital shares. If the rest of the materials used in production are homogeneous, the adjusted capital shares would differ from the original shares, and the adjustment is theoretically exact. In general, even with this adjustment, we are abstracting from the effects of mean reversion in capital stocks in other industries. However, one would expect the omission of these effects to bias our coefficients toward zero. Given the statistical significance of our results presented in columns two and four of Table 1, we believe that this does not undermine our empirical strategy.[5] The omission of scale dependence in other sectors may account for some of the unexplained variation in growth rates. Variation across sectors in other parameters of the model, such as the variance and persistence of productivity shocks, as well as in the depreciation parameter, may account for some of the unexplained variation too.

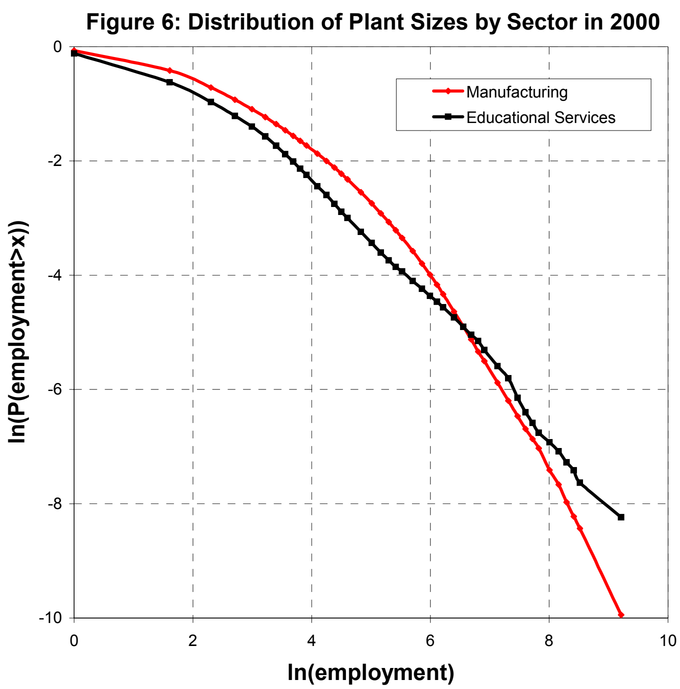
Size Distribution of Firms

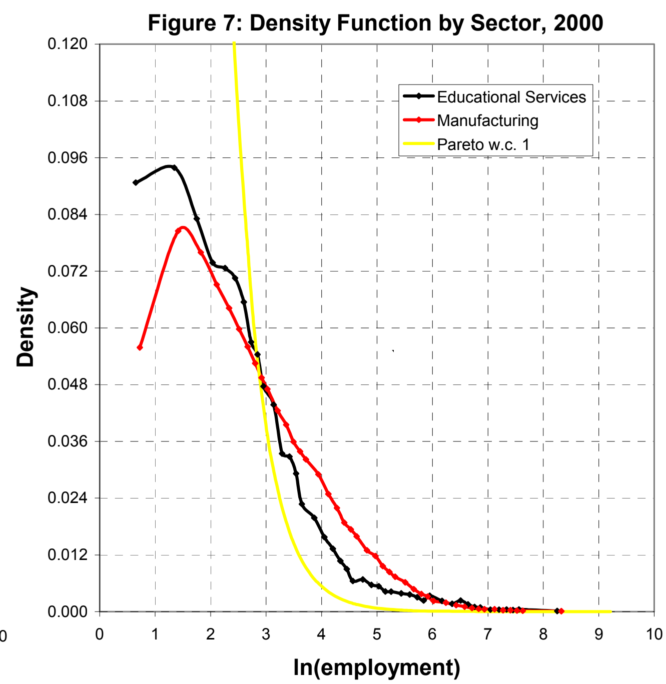
We next turn to the implication of our theory for the size distribution of firms. From the available data we can calculate the share of firms in sector j with employment larger than x_j, which we denote by P_j. If the distribution of firm sizes is Pareto with coefficient one, or growth rates are scale independent, the relationship between $\ln P_j$ and $\ln x_j$ should be linear with slope minus one. If growth rates depend negatively on scale, the tails of the distribution are thinner than the tails of a Pareto with coefficient one, and the relationship is concave. Our theory states that the degree of concavity should be positively related with capital shares for $\alpha_j < 1/2$, and negatively related for capital shares larger than a threshold greater than 1/2. A first look at the data

[5] Adjusting the capital shares increases the number of sectors in our sample with capital shares below one-half from 44 to 52.

is presented in Figures 6 where we plot $\ln P_j$ and x_j for educational services and manufacturing.

The theory predicts that the relationship between $\ln P_j$ and x_j should be more concave for the manufacturing sector. This is indeed the case as can be verified in Figure 6. This representation of the size distribution emphasizes the degree of concavity and makes differences between two distributions particularly clear for large firm sizes. The differences between the distribution are also clear if we look at the density functions.





The density of firm sizes in these two sectors (with normalized means) is presented in Figure 7. It is clear how the distribution of firm sizes in the educational sector has more mass for very small and large firms, and less mass for intermediate firms than in the manufacturing sector. This is particularly clear for small firms in the graph. The figure also compares these distributions with the Pareto distribution with coefficient one (that corresponds to a straight line with slope -1 in Figure 6). The

Pareto distribution with coefficient one has even more mass at the tails and less at the center, consistent with Proposition 6. Both industries have thinner tails than the benchmark, but as the theory predicts, the difference is larger for the manufacturing sector. As emphasized in the introduction, the differences between these distributions are economically large. If the manufacturing sector had the same distribution as the educational sector, around 20% of the labor force in the sector that currently works in medium size firms would need to be reallocated to firms with less than 50 or more than 1000 employees.

In order to test the relationship between capital shares and the size distribution of firms *for all sectors*, we use our new data set on the size distributions of establishments for 1990 and 2000. Given that the BEA has not released data on labor shares at the NAICS three digit industry classification system for 2000, we focus on the results for 1990. We then replicate the study for 2000 with capital shares calculated using the conversion tables for SIC and NAICS classifications provided by the BEA.

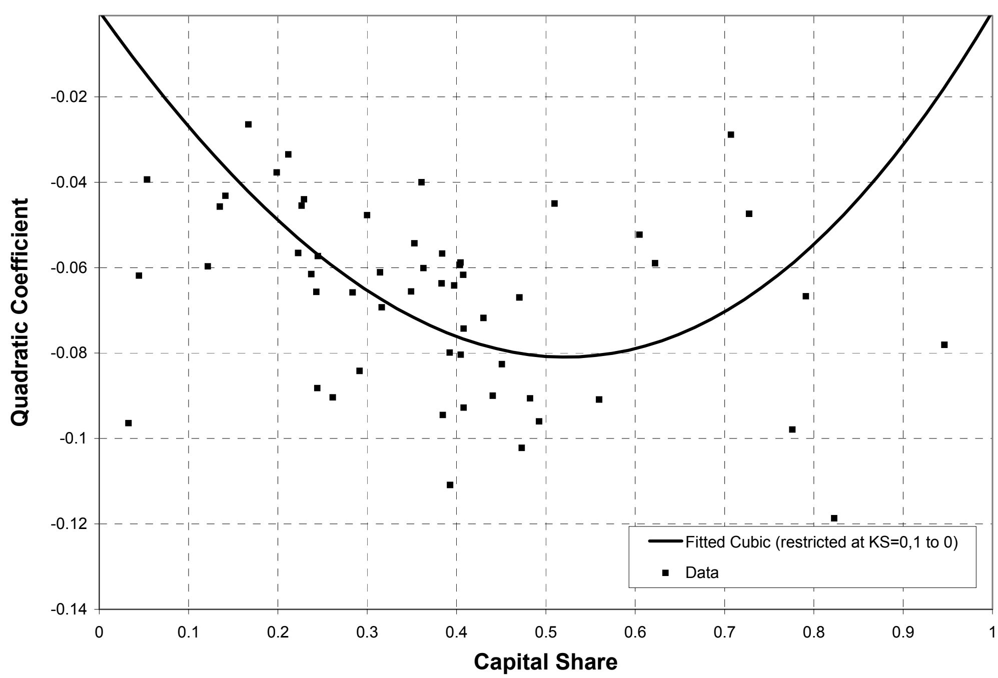
We first estimate the degree of concavity in the relationship between $\ln P_j$ and x_j and then relate those estimates to capital intensity. Towards this, we run the following regression for 1990,

$$\ln P_j = a_j + b_j \ln x_j + c_j (\ln x_j)^2 + \varepsilon_{tj},$$

for each sector using OLS. This gives us estimates of the degree of concavity by sector, c_j. Figure 8 presents a scatter plot of the results with their corresponding capital shares for all the sectors in our sample.

As the theory predicts, there is a clear negative relationship between our estimates of c_j and capital shares for sectors with $\alpha_j < 1/2$. The correlation between α_j and c_j is -0.47. Since all the c_j's are negative, the result implies a positive relationship between the degree of concavity and capital shares, for $\alpha_j < 1/2$. There is some evidence that the relationship is reversed for large capital shares, as the theory also predicts. However, there are not enough sectors with capital shares larger than one half to be able to establish this with statistical confidence. Figure 8 also presents a cubic polynomial, restricted to be zero at capital shares equal to zero and one, that summarizes the non-monotonic relationship in the theory. The fitted cubic shows how the point estimate of the minimum of this relationship is slightly larger than one half. There are two sectors that have capital shares larger than 0.8: Real Estate and the non-farm portion of Agriculture. Both of them use land intensively so the calculated capital share is probably too large.

Figure 8: Quadratic Coefficients and Capital Shares



Once again, the most powerful prediction of the theory is that the estimated c_j's should be negatively related to capital shares for sectors with $\alpha_j < 1/2$. To examine this, we estimate the following regression

$$\ln P_j = \hat{a}_j + \hat{b}_j \ln x_j + \hat{d} (\ln x_j)^2 + \hat{e} \alpha_j (\ln x_j)^2 + \hat{\varepsilon}_{tj},$$

where $\hat{a}_j$ and $\hat{b}_j$ are industry specific coefficients. This amounts to constraining the quadratic term to vary linearly with the capital share. The model now predicts that $\hat{e}$ should be negative and significant for $\alpha_j < 1/2$. The results are in Table 2.

The estimate of $\hat{e}$ for 1990, in the first column of Table 2, is negative and very strongly significant. We also estimated the same regression using NAICS three digit industries in 2000. The capital shares used in this regression are not as trustworthy as the ones in 1990 given that we had to convert the data available for 2000 from the BEA to this industry classification system. Even with this problem, the results presented in the second column of Table 2 show that the estimate of $\hat{e}$ is smaller in absolute value but still negative and strongly significant. The results with adjusted capital shares are presented in the third column of Table 2, which further confirm the empirical significance of the mechanism in our theory.

Table 2

	1990	2000	1999 (adjusted)
$\hat{e}$	-0.0776	-0.0352	-0.058
Standard error	0.0069	0.0019	0.0074
P-value	1.54×10^{-28}	7.80×10^{-07}	8.08×10^{-15}

Exit Rates

Our mechanism, which emphasizes mean reversion in stocks of specific factors, when combined with particular assumptions on preferences, also implies that exit rates should decline with firm size. Furthermore, the rate of decline should vary with capital shares. Figure 9 illustrates this using BITS data for US manufacturing and educational services in 1995-1996. The thick lines represent exit rates in 1995-1996 by establishment size category. The thin solid lines represent the exit rates in 1995 by 1994 size category, and the dashed line by 1992 size class. The graph also presents the linear trends for both sectors, calculated from the exit rates in 1995-1996. The number of firm deaths is divided by the number of surviving firms to compute exit rates.

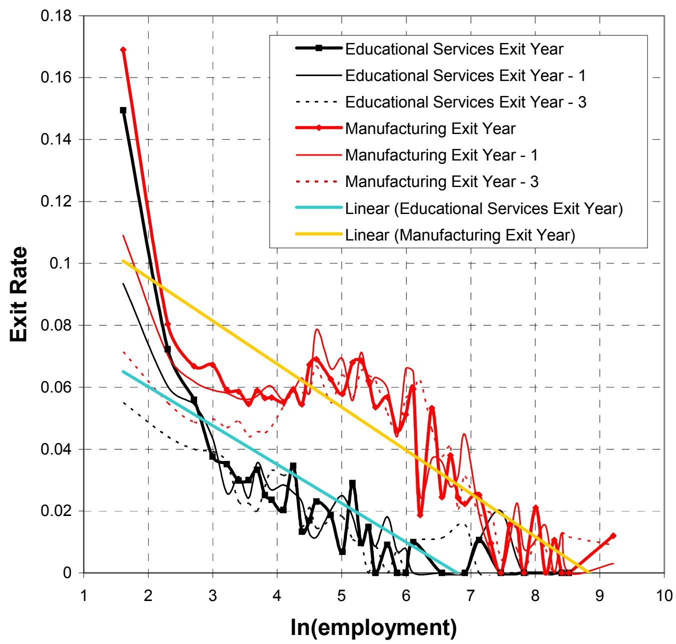


Figure 9: Exit Rates by Sector, 1995-1996

For firms with more than 50 employees the theory does well. Exit rates decline clearly faster with size for manufacturing than for educational services. Overall, the linear trend in manufacturing is steeper than in educational services, although the difference is small given the large variance (a slope of -0.0139 for manufacturing and -0.0126 for educational services).[6]

A problem in testing this implication is that small firms face adjustment costs. Some firms may downsize progressively over a couple of years before exiting, and new

[6] Orr (1974), Gorecki (1976), Hause and Du Rietz (1984) and MacDonald (1986) found that firm exit rates were negatively related to measures of capital intensity by industry. Given that these studies do not distinguish among firms with different sizes, the negative relationship may be the result of the dependence predicted by our theory. This would be the case if firms in capital intensive sectors are larger on average. The evidence is, however, weak and hopefully future research, with suitable data, will be able to test our prediction.

entrants may enter at a suboptimal size and then grow to their chosen entry size. This implies that we want to track a firm for several years before they exit and several years after they enter. In fact, we can see in Figure 9 that firms with less than 50 employees tend to start downsizing several years before they exit. This is, however, not the case for firms with more than 50 employees. Hence, selection, as emphasized by other theories, may be important for very small firms but does not seem to be as important for the exit pattern of large firms.[7] In the introduction we discussed several theories that have explained some of the facts we address in this paper by adding particular frictions that small firms face. Our theory does not include any of these frictions nor does it address divisibilities in technology. Adding these elements could help explain why the theory does not do as well in matching the qualitative relationship in Figure 6 for firms with less than 50 employees, and cannot match the data on entry.[8]

CONCLUSION

In this paper we have constructed a theory that is consistent with some well known facts on firm dynamics and firm size distributions. The mechanism emphasizes the role of accumulation of industry specific factors. We have shown that this mechanism is robust to institutional and economic differences across sectors and countries. We claim that the ubiquitous presence of these facts has to be the result of a mechanism that is present in a variety of circumstances. The central role of accumulation of specific factors in the theory led us to think about cross sectoral differences in the importance of these specific factors in production, and in particular capital intensity.

[7] In these theories large firm exit only if it is possible for them to receive very large negative shocks.

[8] If we replicate Figure 9 for entry rates, we find that entry rates are decreasing with size in manufacturing and educational services. This is a contradiction of our theory's prediction that entry rates should increase with size. We believe this is due to frictions (e.g. financial) that constrain firm size at entry.

The model has a striking prediction of how growth rates and size distributions of firms should vary as we change capital intensity. When capital intensity is low, an increase in the importance of specific factors leads to more scale dependence in growth rates and a distribution of firm sizes with thinner tails. This implication is reversed for large capital intensities. Together, this non-monotone pattern provides a novel and unexamined prediction of our theory. Since it was the theory that guided our focus on this particular dimension of the data, the available evidence in the empirical literature is only indirect. Consequently, we take this prediction to the data and show that it is a surprisingly good description for a cross-section of US sectors.

Our theory implies that exit rates should decline with size. Conversely, it implies that entry rates should increase with size. The model's implications on exit rates are consistent with the empirical evidence. However, entry rates do not seem to increase with size; new entrants start their businesses at a small scale. On the one hand, it is puzzling that a theory that does a good job in explaining many related phenomena is not successful in this particular dimension. On the other, we built a theory under the strong assumption that entry and exit is frictionless. This assumption is particularly important for entrants, especially if we look at their size in their first year of existence. Detailed longitudinal data on entry and exit may shed light on whether looking at size several years after entry eliminates this mismatch. It is easy to think about particular frictions faced by new entrants.

In the introduction we commented on different studies that have emphasized financial as well as other types of frictions. What we show in this paper is that even though these frictions are important for entry, they are not needed to generate any of the other empirical observations. This points to frictions in entry that might be alleviated with particular policies. It is important, however, that these policies do not interfere with the growth and exit of existing firms; processes that are well described by our efficient economy. We have shown that our results are not sensitive to policies that affect firms independently of their size. Restuccia and Rogerson (2004) have argued that scale dependent policies may have large effects on efficiency and these types of policies may also affect some of our implications. International evidence on firms dynamics and the size distribution of firms, when combined with our benchmark, could shed some light on the empirical significance of scale dependent policies.

By emphasizing the accumulation of specific factors, our theory also makes prediction for the future evolution of the firm size distribution. The ongoing specialization of developed economies in services will have important consequences on firm sizes and firm dynamics. Our theory predicts that this will lead to a more dispersed distribution of firm sizes, where we will see more small and more very large firms. Similarly, the development and spread of general purpose technologies, such as computers, will increase the dispersion of firm sizes in all sectors. This effect will be particularly important in sectors that heavily substitute specific factors with general purpose capital. These arguments suggest that we are moving towards an economy in which the dominance of large firms in some industries, like Walmart, will coexist increasingly with large numbers of small firms in different industries within the same sector, like bakeries or tailors. This trend is the natural result of the efficient division of an industry's production among firms.

REFERENCES

[1] Albuquerque, R. and H. Hopenhayn (2002) "Optimal Lending Contracts and Firm Dynamics." University of Rochester - Center for Economic Research Working Paper 493.

[2] Audretsch, D.B., L. Klomp, and A.R. Thurik (2002) "Gibrat's Law: Are the Services Different?" ERIM Report Series 2002-04.

[3] Axtell, R. L. (2001) "Zipf distribution of US firm sizes", *Science*, 293, 1818-1820.

[4] Aw, B., S. Chung, and M. Roberts. (2003) "Productivity, Output and Failure: A Comparison of Taiwanese and Korean Manufacturers." working paper, Penn State University.

[5] Caves, R. E. (1998). "Industrial Organization: New Findings on the Turnover and Mobility of Firms." *Journal of Economic Literature*, 36, 1947-1983.

[6] Clementi, G. and H. Hopenhayn (2002) "A Theory of Financing Constraints and Firm Dynamics." University of Rochester - Center for Economic Research Working Paper 492.

[7] Dunne, T., M. Roberts, and L. Samuelson, 1989. "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector." *Journal of Labor Economics*, Vol. 7 (1) pp. 48-71.

[8] Dunne, T., M. Roberts, and L. Samuelson, 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics*, Vol. 104 (4) pp. 671-98.

[9] Ericson, R. and A. Pakes (1995). "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1): 53-82.

[10] Evans, D. S. (1987a). "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy,* 95(4): 657-674.

[11] Evans, D. S. (1987b). "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries." *Journal of Industrial Economics,* 35(4): 567-81.

[12] Garicano, L. and E. Rossi-Hansberg (2004). "Inequality and the Organization of Knowledge." *American Economic Review P&P*, forthcoming.

[13] Geroski, P. (1995). "What do we Know About Entry?" *International Journal of Industrial Organization*, 13, 421-440.

[14] Geroski, P. (1983). "The Empirical Analysis of Entry: A Survey." University of Southampton, mimeo.

[15] Geroski, P. and J. Schwalbach (1991). *Entry and Market Contestability.* Oxford, Blackwell.

[16] Gibrat, R. (1931). *Les inégalités économiques*, Libraire du Recueil Sirey.

[17] Gorecki, P. K. (1976). "The Determinants of Entry by Domestic and Foreign Enterprises in Canadian Manufacturing." *Review of Economics and Statistics*, 58: 485-488.

[18] Hall, B. H. (1987). "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector." *Journal of Industrial Economics*, 35(4): 583-606.

[19] Hause, J. C. and G. Du Rietz (1984). "Entry, Industry Growth, and the Microdynamics of Industry Supply." *Journal of Political Economy*, 92: 733-757.

[20] Hopenhayn, H. A. (1992). "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(2): 1127-1150.

[21] MacDonald, J. M. (1986). "Entry and Exit on the Competitive Fringe." *Southern Economic Journal*, 52: 640-658.

[22] Mansfield, E. (1962). "Entry, Gibrat's Law, Innovation, and the Growth of Firms." *American Economic Review*, 52(5): 1023-51.

[23] Orr, D. (1974). "The Determinants of Entry: A Study pf the Canadian Manufacturing Industries." *Review of Economics and Statistics*, 56: 58-66.

[24] Restuccia, D. and R. Rogerson (2004). "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." Unpublished paper, Arizona State University.

[25] Rossi-Hansberg, E. and M. L. J. Wright (2004). "Urban Structure and Growth." Unpublished paper, Stanford University.

[26] Scherer, F. M. (1980). *Industrial Market Structure and Economic Performance*. Boston, Houghton Mifflin.

[27] Schmalensee, R. (1989). "Inter-Industry Studies of Structure and Performance." *Handbook of Industrial Organization*. R. Schmalensee and R. Willig. Amsterdam, North Holland. 2: 951-1009.

[28] Sutton, J. (1997). "Gibrat's Legacy." *Journal of Economic Literature*, 35(1): 40-59.

[29] van Ark, B. and E. Monnikhof (1996). "Size Distribution of Output and Employment: A Data Set for Manufacturing Industries in Five OECD Countries, 1960s-1990." OECD Economics Department Working Paper 166.