

UC Santa Cruz

Archived UCSC Economics Department Seminars

Title

Firm Size Dynamics in the Aggregate Economy

Permalink

<https://escholarship.org/uc/item/4rs4202s>

Author

Wright, Mark

Publication Date

2004-12-12

Firm Size Dynamics in the Aggregate Economy*

Esteban Rossi-Hansberg

Mark L.J. Wright

Stanford University

Stanford University

December 12, 2004

Abstract

Why do firm growth and exit rates decline with size? What determines the size distribution of firms? This paper presents a theory of firm dynamics that simultaneously rationalizes the basic facts on firm growth, exit, and size distributions. The theory emphasizes the accumulation of industry specific human capital in response to industry specific productivity shocks. The theory implies that firm growth and exit rates should decline faster with size, and the size distribution should have thinner tails, in sectors that use human capital less intensively, or correspondingly, physical capital more intensively. In line with the theory, we document substantial sectoral heterogeneity in US firm dynamics and firm size distributions, which is well explained by variation in physical capital intensities.

1. 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. In this paper we propose an explanation of this scale dependence that relies on the response of production

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

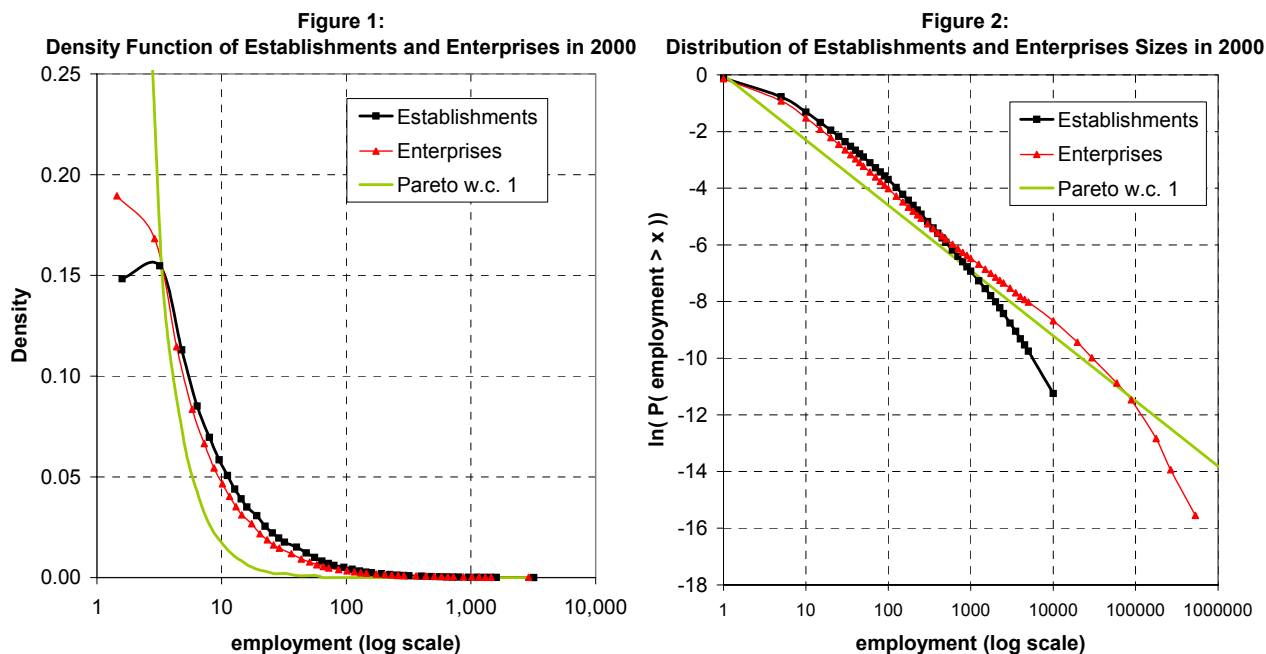
decisions to the allocation and accumulation of industry specific human capital. Our theory can simultaneously rationalize the facts on growth and exit rates as well as the size distribution of firms. In addition, the theory implies that differences in the importance of industry specific human capital, and therefore also physical capital, across sectors should lead to cross-sectoral variation in the degree of scale dependence within a sector. We present evidence from a new data-set to document these facts for the US economy. We find that, as implied by our theory, US sectors with larger physical capital shares exhibit significantly more scale dependence in firms dynamics and size distributions.

A large literature beginning with Gibrat (1931) has examined the size distribution of firms. Figure 1 illustrates what we mean by scale dependence in the size distribution of firms by comparing the densities of establishment sizes (employment at operations at a single location) and enterprises (employment at operations under common ownership or control) for the US economy in 2000 to a commonly used benchmark: a Pareto distribution with shape coefficient one (see, for example, Axtell (2001)). The Pareto distribution is scale independent in the sense that the distribution is invariant to truncation of the left tail. The figure shows that the enterprise and establishment size distributions are similar, reflecting the fact that only the very largest enterprises possess more than a single establishment. Importantly, both distributions have thinner tails than the Pareto benchmark: there is scale dependence in the size distribution of firms. In Figure 2, we present these data in a different format in order to emphasize the right tail of the distribution. If production units are distributed according to a Pareto distribution, the logarithm of the share of production units greater than a particular employment size varies linearly with the 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.¹

We interpret the similarity between both curves in Figure 2 as evidence that the

¹In Figure 2 one can see that the distributions of enterprises and establishments are similar for units with less than 400 employees reflecting the fact that most enterprises are formed by one establishment. The curve for establishments is clearly concave, as is the one for enterprises although at a larger scale. The latter finding is surprising in light of the commonly held view that the distribution of enterprises is well described by a Pareto distribution with coefficient one.

same economic forces determine the size of establishment and enterprises. In Figure 2 we include data for enterprises with close to one million employees to highlight the previous statement. Hence, in what follows, we suppress the distinction and refer to production units simply as firms. Nevertheless, the theory we develop below refers to the technology of a single production unit and does not address questions of ownership or control. Consequently throughout the paper we focus solely on establishment data.

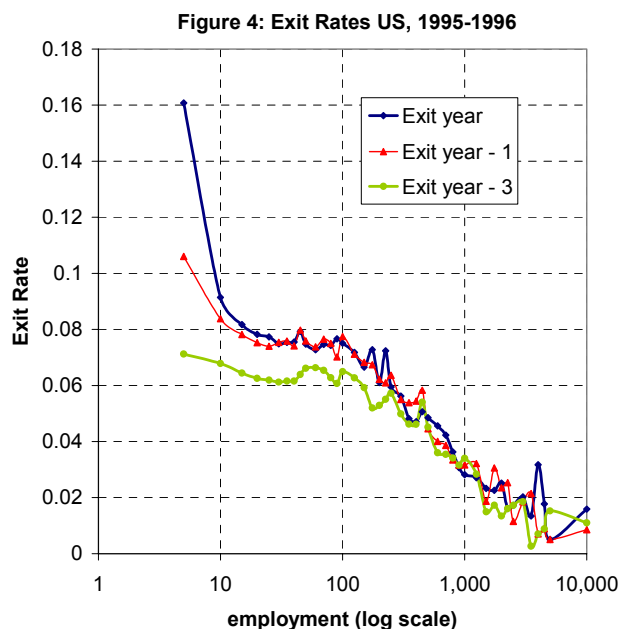
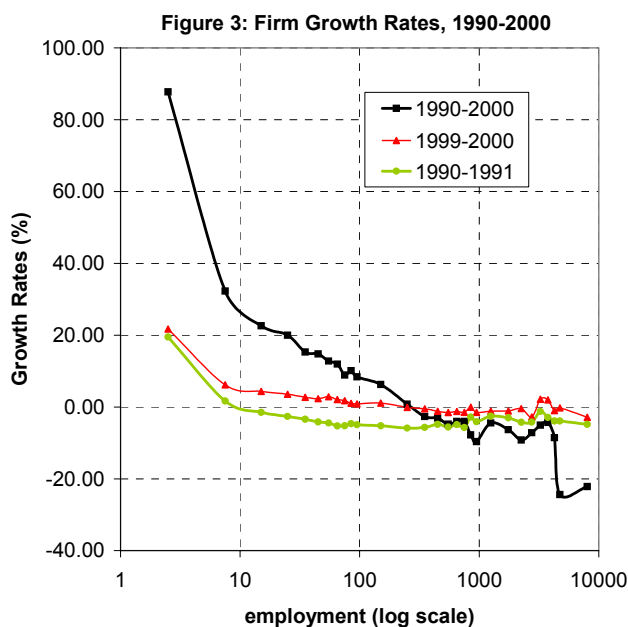


Firm dynamics are also scale dependent. For example, it is well-known that small firms grow faster than large firms, at least when attention is restricted to those firms that remain in operation.² This is illustrated in Figure 3 which plots growth rates by firm size for the US over both one and ten year 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.

²This fact was most forcefully demonstrated by Mansfield (1962) in his study of firms in the steel, 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. See also the surveys by Scherer 1980, Geroski 1995, Sutton 1997, and Caves 1998, who document the robustness of these results across time, industries and countries.

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 scale dependence in exit rates: small firms are more likely to exit than large firms. 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. Several features in this figure should be noted. First, exit rates decline substantially with size, even for firms with more than 1000 employees. Second, there is no evidence of the “Shadow of Death”: firms declining in size in the years leading up to their death (Griliches and Regev 1995). There is, however, strong evidence that recent entrants have higher exit rates as illustrated by the increased mass of small firms as they approach their exit date. This suggests that selection is important for small young firms, but not for medium and large ones.



Figures 1 through 4 illustrate the scale dependence in firm dynamics and size distributions for the US over the 1990s. However, scale dependence has also been documented over different time periods, sectors, and countries. This is surprising given

the enormous diversity of institutions, market structures, and technology. The robustness of these facts demands a theory that emphasizes forces common to a variety of circumstances and sectors. Moreover, it requires a theory where these facts survive aggregation and are consistent with aggregate evidence.

To address these facts, we propose an aggregate theory of firm dynamics based on the accumulation of industry specific human capital. We present a stochastic growth model with multiple goods. The set of goods in the economy is divided into subgroups that we call *sectors*. Each sector is in turn formed by a collection of goods that we call *industries*. Firms operate in only one industry and hire labor and industry specific human and physical capital. As long as technology exhibits diminishing returns to human capital at the firm level, and this is preserved by aggregation within an industry, an abundance of human capital leads to low rates of return and slower accumulation of human capital. Conversely, if the stock of the human capital is relatively low, rates of return are high and accumulation is fast. This process, which is at the heart of the resource allocation mechanism in the economy, leads to mean reversion in the stock of industry specific human capital. As long as firms respond monotonically to fluctuations in factor prices driven by the stock of human capital, mean reversion in these stocks leads to mean reversion in firm sizes. This results in small firms growing faster than large firms.

The same process also implies that exit rates decline with size. To see this, note that, 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 human capital. As a result, the mechanism is robust to a variety of different environ-

ments. To establish this, we also consider different production technologies, within industry firm heterogeneity, alternative mechanisms for the accumulation of human capital such as learning by doing, and differences in the form of product market competition.

The emphasis on the accumulation and allocation of specific human capital implies that firm growth and exit rates should decline faster with size in sectors that use human capital less intensively. In turn, this implies that the tails of the size distribution of firms should be thinner the smaller the human capital share. The rate of accumulation of industry specific human capital is tied to industry production either because the same factors of production are used to generate new industry specific knowledge, or because past production affect the stock of this knowledge directly through learning by doing. The elasticity of factor prices to factor stocks is positively related to the share of the factor in production. These prices in turn determine the accumulation of industry specific factors and therefore the degree of mean reversion. Hence, the degree of mean reversion decreases with human capital intensity, just as in the neo-classical growth model the speed of convergence decreases with the physical capital share. Unlike human capital, physical capital investments are tied to production in a wide variety of sectors that diffuses this mechanism. We show that the process of entry and exit of firms ensures that industry production will display constant returns to scale, and so physical capital intensities are negatively related to human capital intensities. This implies that the intensity of physical capital in production is positively related to the degree of mean reversion in human capital and, hence, to the degree of mean reversion in firm sizes.

We assess the relationship between capital shares and firm scale dependence using a new data-set 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 sectors. 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 physical capital shares. We then proceed to show that this same relationship is reflected in exit rates and in significant differences in the size distribution of firms across sectors. The differences are large. For example, in order to make the size distribution of firms in the physical capital intensive manufacturing

sector 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 firms 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

In contrast to our approach, most recent theoretical attempts to explain the size distribution of firms have focused on particular dimensions of the dynamics of firms in an industry assuming elastically supplied factors of production. 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), Ericson and Pakes (1995) and Luttmer (2004) 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

³Relatively 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 physical capital intensity by industry.

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; 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 human capital in an industry. Many of the mechanisms in the literature undoubtedly contribute towards an explanation of firm dynamics. This paper shows, we believe, that the accumulation of industry specific human capital 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, along with a discussion of this link between our theory and the empirical work on specific human capital by Kambourov and Manovskii (2002). Section 4 describes our data, and presents results that show that firm growth and exit rates, as well as the firm size distribution, vary with physical capital shares in precisely the way predicted by our theory. Section 5 concludes.

2. THE MODEL

We present a stochastic dynamic aggregate model in which firms are perfectly competitive. Labor is mobile across all industries, while both physical and human capital are 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.

2.1 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 do not value leisure and order their preferences over state contingent consumption streams $\{C_t\}$ of the single final good according to

$$(1 - \delta)E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \ln \left(\frac{C_t}{N_t} \right) \right], \quad (1)$$

where δ is the discount factor of the household, 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 household's consumption bundle, and that the household as a whole sums the valuations of each of its members.

The household produces the final good by combining quantities of J different intermediate goods $\{Q_{tj}\}$ according to the constant returns to scale production function

$$C_t + \sum_{j=1}^J X_{tj} = B \prod_{j=1}^J (Q_{tj})^{\theta_j}. \quad (2)$$

The final good can be used for consumption, as well as for investment in physical capital in each of the J intermediate good industries X_{tj} . We distinguish these intermediates by what we refer to as a *sector* and an *industry*. In particular, we assume that there are S sectors in this economy, and that each sector contains J_s industries, where $s = 1, \dots, S$. Each industry produces a single distinct good so that there are $J = \sum_{s=1}^S J_s$ goods being produced in this economy. Sectors differ according to the methods by which output is produced and factors are accumulated; within a sector, the parameters governing production and accumulation of factors for each industry are the same. We also assume that each industry within a sector has the same share in production of the final good so that $\theta_j = \theta_i$ for all i, j in sector s . Importantly, each industry within a sector receives its own productivity shock and accumulates its own stocks of human and physical capital. This is important below: because each industry within a sector evolves separately, according to a process governed by the

same parameters, we will be able to characterize the invariant distribution of firm sizes within each sector. In thinking about the data, we define our sectors to be roughly comparable to the list of 3 digit NAICS classifications, while our industries map into NAICS industries at a much finer level of disaggregation.

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

$$\sum_{j=1}^J N_{tj} \leq N_t. \quad (3)$$

Households also rent out their stocks of each of the J industry-specific physical and human capital stocks, which we denote by K_{tj} and H_{tj} respectively. Physical capital accumulates according to the log-linear form

$$K_{t+1j} = K_{tj}^{\lambda_j} X_{tj}^{1-\lambda_j}. \quad (4)$$

This log-linear form for physical capital accumulation has grown increasingly popular as a device for modelling adjustment of physical capital while still admitting closed form solutions. Here λ_j captures the importance of past physical capital stocks to the amount of capital next period: if λ_j is one, capital does not evolve and is a fixed factor; if λ_j is zero, physical capital depreciates fully each period.

Human capital is also assumed to accumulate according to a log-linear function

$$H_{t+1j} = A_{t+1j} H_{tj}^{\omega_j} I_{tj}^{1-\omega_j}.$$

Here, A_{t+1j} is an industry specific shock that is assumed to be i.i.d. with compact support $[\underline{A}_j, \bar{A}_j]$ and is designed to capture the random accumulation within an industry, while I_{tj} is an investment in human capital accumulation. These industry specific productivity shocks are the only source of randomness in our model.

We assume that I_{tj} is denominated in terms of the output of the industry itself, in order to capture the idea that industry specific learning requires some industry specific inputs, so that the resource constraint for output of industry j , Y_{tj} , is

$$Q_{tj} + I_{tj} = Y_{tj}.$$

In our framework there are no externalities: Human capital investments are paid by households, and they rent the new human capital for use in production. In Section 3, below, we also present an extension of the model which allows for learning-by-doing externalities and show that it has similar properties. Moreover, with learning-by-doing externalities, households do not appropriate the rewards to industry-specific learning, which is consistent with empirical evidence on industry specific human capital (for example, Kambourov and Manovskii 2002). The assumption that human capital accumulation responds to industry-specific production levels is essential for our results as it will serve as the primary source of industry specific mean reversion.

Finally, as noted above, we assume that the accumulation parameters are identical across all industries within a sector; that is, $\omega_j = \omega_i$ and $\lambda_j = \lambda_i$ for all i, j in sector s . The household begins with initial stocks of these specific factors denoted by K_{0j} and H_{0j} .

2.2 Firms

Production within each industry takes place in production units that we call firms. To begin, for simplicity, we abstract from firm specific heterogeneity and assume that each firm in industry j at time t has access to the same production technology; we relax this assumption in Section 3 below. To produce in a period, the firm must pay a fixed cost F_j that period. Once the fixed cost has been paid, the firm hires industry- j -specific physical capital k_{tj} , in combination with an industry- j -specific labor input that is, in turn, produced by combining raw labor n_{tj} with industry- j -specific human capital, h_{tj} , and produces according to

$$y_{tj} = \left[k_{tj}^{\alpha_j} \left(h_{tj}^{\beta_j} n_{tj}^{1-\beta_j} \right)^{1-\alpha_j} \right]^{\gamma_j}. \quad (5)$$

Here $\gamma_j < 1$ captures the extent of decreasing returns to production which, in combination with the fixed cost, ensures that average costs are “U-shaped” and serves to pin down the size of the firm. The parameter α_j governs the share of physical capital in value added, while β_j captures the share of human capital in the labor aggregate. Both production parameters and the process governing evolution of the productivity shock are assumed to be common across all industries within a sector: $\alpha_j = \alpha_i$, $\beta_j = \beta_i$ and $\gamma_j = \gamma_i$ for all i, j in sector s .

None of our results depend upon the denomination of the fixed cost, and so to begin we assume that it is denominated in the units of the firms output. This has the expositional advantage of pinning down the scale of production of the plant (measured in terms of output), so that we can easily analyze the effects of changes in factor prices on the size of the firm (measured in terms of the number of employees); we return to this assumption below.

2.3 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, this 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 $\{C_{tj}, X_{tj}, I_{tj}, N_{tj}, \mu_{tj}, H_{tj}, K_{tj}\}_{t=0, j=1}^{\infty, J}$ so as to maximize household welfare

$$(1 - \delta)E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \ln \left(\frac{C_t}{N_t} \right) \right], \quad (6)$$

subject to the resource constraint on the final good

$$C_t + \sum_{j=1}^J X_{tj} = B \prod_{j=1}^J (Y_{tj} - I_{tj})^{\theta_j}, \quad (7)$$

for all dates and states, the resource constraint on each intermediate good

$$Y_{tj} = \left[K_{tj}^{-\alpha_j} \left(H_{tj}^{\beta_j} N_{tj}^{1-\beta_j} \right)^{1-\alpha_j} \right]^{\gamma_j} \mu_{tj}^{1-\gamma_j} - F_j \mu_{tj}, \quad (8)$$

for each industry, date and state, the accumulation equations for each industry-specific factor

$$K_{t+1j} = K_{tj}^{\lambda_j} X_{tj}^{1-\lambda_j}, \quad (9)$$

and

$$H_{t+1j} = H_{tj}^{\omega_j} I_{tj}^{1-\omega_j}, \quad (10)$$

for all industries, dates and states, and the constraint on labor allocation

$$N_t = \sum_{j=1}^J N_{tj}, \quad (11)$$

for all dates and states.

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

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

which implies

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

This leads to an equilibrium firm size that depends on the amount of factors in the industry according to

$$n_{tj} = \frac{N_{tj}}{\mu_{tj}} = \left[\frac{F_j}{1 - \gamma_j} \right]^{\frac{1}{\gamma_j}} \left(\frac{N_{tj}}{K_{tj}} \right)^{\alpha_j} \left(\frac{N_{tj}}{H_{tj}} \right)^{\beta_j(1-\alpha_j)}. \quad (12)$$

If the stock of specific factors is high relative to the amount of labor employed in the industry (which corresponds to the case of relatively cheap specific factor prices), firms size measured in terms of the number of employees will be small. Similarly, mean reversion in the stock of relative specific factor stocks will drive mean reversion in firm sizes. Importantly, the qualitative nature of the relationship between factor stocks and firm size can be reversed, without changing the result that mean reversion in these stocks produces mean reversion in firms sizes. In the next section, we show that the incentive to accumulate produces precisely the required mean reversion in the general equilibrium of our model.

Substituting for the optimal number of firms into the resource constraint gives

$$Q_{tj} + I_{tj} \leq \gamma_j \left[\frac{1 - \gamma_j}{F_j} \right]^{\frac{1-\gamma_j}{\gamma_j}} K_{tj}^{\alpha_j} \left(H_{tj}^{\beta_j} N_{tj}^{1-\beta_j} \right)^{1-\alpha_j}.$$

This is our first main result: by varying the number of firms, each of which produces at the bottom of its average cost curve, the industry behaves as though it has constant returns to scale.

The result is an entirely standard log-linear multi-sector growth model with a new constant returns to scale production function.⁴ As a result of the log-linear assumptions, we get the well-known result (see, for example, the appendix to Rossi-Hansberg and Wright (2004a)) that income and substitution effects offset to ensure that a fixed proportion of the labor supply is allocated to each industry, a fixed proportion of the final good is consumed, while fixed proportions are invested in each industry, and a fixed proportion of the output of each intermediate input is used for investment in human capital specific to that industry.

2.4 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 (12) we find that the growth rate of a firm in industry j is given by

$$\begin{aligned} \ln n_{t+1j} - \ln n_{tj} &= (\alpha_j + \beta_j (1 - \alpha_j)) g_N - \alpha_j [\ln K_{t+1j} - \ln K_{tj}] \\ &\quad - \beta_j (1 - \alpha_j) [\ln H_{t+1j} - \ln H_{tj}], \end{aligned}$$

and substituting for the evolution of human capital we get

$$\begin{aligned} \ln n_{t+1j} - \ln n_{tj} &= (\alpha_j + \beta_j (1 - \alpha_j)) g_N - \alpha_j [\ln K_{t+1j} - \ln K_{tj}] \\ &\quad - \beta_j (1 - \alpha_j) [\ln A_{t+1j} + (\omega_j - 1) \ln H_{tj} + (1 - \omega_j) I_{tj}]. \end{aligned}$$

This equation reveals that the growth rate of a firm in industry j is driven by three factors. The first is the deterministic growth in the aggregate labor supply g_N which, other things equal, encourages firms to expand in size over time. We will often assume that either population growth is zero, or that firms growth rates are being measured relative to trend, in order to abstract from this term. The second factor is the growth in industry specific physical capital. However, as physical capital

⁴In a related paper Jones (2004) shows how a Pareto size distribution of firms leads to an aggregate Cobb-Douglas production function.

investment in each industry is a constant proportion of the aggregate production of the final good, this is also determined by aggregate forces. Over time, if the number of industries is large so that industry-specific randomness washes out in the aggregate, the aggregate economy converges to a steady state and this term will be a constant. In what follows we assume this is the case in order to focus on industry specific variation; in general, the results that follow can be thought of as being conditioned upon the state of the aggregate economy. Finally, we have the contribution of industry specific variability, which works through the shock to human capital accumulation, and the level of industry output which affects human capital accumulation through I_{tj} : if industry output is high, human capital accumulation proceeds, on average, at a faster pace.

Before turning to a discussion of scale *dependence* in growth rates, it is useful to begin by examining the conditions under which we get scale *independent* growth or, in other words, the conditions under which we get Gibrat's Law. First, suppose we eliminate human capital as a factor of production by either reducing the importance of labor as a whole (that is, reducing $(1 - \alpha_j)$) or reducing the importance of human capital in producing labor services (that is, reducing β_j). In this case, the firm grows at a deterministic rate that is independent of scale. This is due to the fact that the only source of industry-specific randomness comes from shock to the accumulation of human capital.⁵ Second, suppose that human capital is accumulated exogenously, or that $\omega_j = 1$: this ensures that output in an industry has no effect on the pace of its human capital accumulation.⁶ With the aggregate economy in steady state, the growth rate of the firm becomes

$$\ln n_{t+1j} - \ln n_{tj} = (\alpha_j + \beta_j(1 - \alpha_j))g_N - \beta_j(1 - \alpha_j)\ln A_{t+1j},$$

which is a constant plus an i.i.d. random variable: the growth rate of the firm is

⁵One way to retain randomness in production while still eliminating human capital as a factor is to scale up the shock to human capital by the inverse of the elasticity of human capital in production $\beta_j(1 - \alpha_j)$. In this case, the growth rate of the firm also satisfies Gibrat's Law and becomes $\ln n_{t+1j} - \ln n_{tj} = \alpha_j g_N - \ln \hat{A}_{t+1j}$, where \hat{A}_{t+1j} is the scaled shock process.

⁶If $\omega_j = 1$, human capital in industry j , and consequently also output, is difference stationary. If industry j is of positive measure, the aggregate physical capital stock will not in general converge to a steady state under this assumption. As long as $1 - \omega_j$ is positive, no matter how small, the existence of a steady state is preserved. When we refer to the case of $\omega_j = 1$ below, we shall think of $1 - \omega_j$ arbitrarily small but positive.

independent of the size of the firm.

To see how firm growth rates depend upon firm size in general, assume as before that population growth is zero and that the aggregate economy is in steady state so that physical capital is constant in all industries. Then using equation (12) the growth rate of the firm, after substituting for I_{tj} , can be written as

$$\ln n_{t+1j} - \ln n_{tj} = n^C - (1 - \omega_j) (1 - \beta_j + \alpha_j \beta_j) \ln n_{tj} - \beta_j (1 - \alpha_j) \ln A_{t+1j}, \quad (13)$$

where n^C is a constant term that depends on the physical capital stock. We summarize the results of this discussion in the following proposition in which we emphasize the effect of changes in physical capital intensity, an observable parameter which we focus upon in our empirical analysis.

Proposition 1 *Firm growth rates are weakly decreasing in size. The higher is the physical capital share, the faster growth rates decline with size. The growth rate of firms is independent of its size only if either human capital is not a factor of production (in the limit as β_j or $(1 - \alpha_j)$ are equal to 0), or human capital evolves exogenously (in the limit as ω_j approaches one).*

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: there is exit whenever firm sizes grow on average. 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 the higher the physical capital share, the faster the exit rate decreases with firm size.

Corollary 2 *Firm exit rates are weakly decreasing in size. The higher is the physical capital share, the faster exit rates decline with size. The exit rate of firms is independent of size only if either human capital is not a factor of production (in the limit as β_j or $(1 - \alpha_j)$ are equal to 0), or human capital evolves exogenously (in the limit as ω_j approaches one).*

These implications for the relationship between physical 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. Away from these limits, when there is mean reversion in firm growth rates, it can be established that there exists a unique invariant distribution that has thinner tails than implied by Zipf's Law: there is a relative absence of very small, and very large, firms. We can also establish that the tails of the size distribution become thinner as physical capital shares increase. These claims are proven in the following three propositions.

Proposition 3 (*Zipf's Law*) *If either human capital is not a factor of production (in the limit as β_j or $(1 - \alpha_j)$ are equal to 0), or human capital evolves exogenously (in the limit as ω_j approaches one), the size distribution of firms converges to a Pareto distribution with shape coefficient one.*

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

Outside of these special cases, 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 which follows directly from our assumption that log productivity levels lie in the compact set $[\ln \underline{A}, \ln \bar{A}]$ for some \underline{A} suitably small and \bar{A} suitably large, and that firm sizes are measured relative to trend (or equivalently that population growth is zero). These assumptions imply that

$$\ln n_{tj} \in LN \equiv \frac{\beta_j (1 - \alpha_j)}{(1 - \omega_j) (1 - \beta_j (1 - \alpha_j))} [-\ln \bar{A}, -\ln \underline{A}].$$

Proposition 4 *For any $\alpha_j, \beta_j, \omega_j \in (0, 1)$, there exists a unique invariant distribution over firm sizes 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

$$\ln n_{t+1} = g(n_t, A_{t+1}) \equiv -\ln A_{t+1} + \left(1 - (1 - \omega_j)(1 - \beta_j(1 - \alpha_j))\right) \ln n_t,$$

where we have assumed that the population size is fixed (alternatively, we could work with variations from trend). This lies in the compact set LN defined above. Let μ be the probability measure over A . Then, the probability of a transition from a point n to a set S is given by

$$Q(n, S) = \mu(A : g(n, A) \in S).$$

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

$$(Tf)(n) = \int_{LN} f(n') Q(n, dn') = \int_{\underline{A}}^{\bar{A}} f(g(n, A)) d\mu(A).$$

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

$$(T^*\lambda)(S) = \int_{LN} Q(n, S) \lambda(dn).$$

Since the set LN 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 $\ln n$, and $\ln A$. Since g is continuous and bounded, if f is continuous and bounded, $f(g(\cdot))$ will be 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 $f : LN \rightarrow \mathbb{R}$ is a non-decreasing function, then so is Tf . But this follows from the fact that the g is non-decreasing in n . Hence $f(g(n, A))$ is non-decreasing in n and therefore so is Tf .

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

$$Q\left(\frac{-\ln \bar{A}\beta_j(1 - \alpha_j)}{(1 - \omega_j)(1 - \beta_j(1 - \alpha_j))}, \left[c, \frac{-\ln \underline{A}\beta_j(1 - \alpha_j)}{(1 - \omega_j)(1 - \beta_j(1 - \alpha_j))}\right]\right) \geq \eta,$$

and

$$Q \left(\frac{-\ln \underline{A} \beta_j (1 - \alpha_j)}{(1 - \omega_j) (1 - \beta_j (1 - \alpha_j))}, \left[\frac{-\ln \bar{A} \beta_j (1 - \alpha_j)}{(1 - \omega_j) (1 - \beta_j (1 - \alpha_j))}, c \right] \right) \geq \eta.$$

Let $c = 0$. As g is continuous and decreasing in A , there exists an A' such that for all $A \leq A'$, $g(n, A) > 0$. Let $\eta' = 1 - \mu(A')$. Similarly there exists an A'' such that for all $\varepsilon \leq A''$, $g(n, A) < 0$. Let $\eta'' = 1 - \mu(A'')$. Call the minimum of these probabilities η . Then $c = 0$ 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^* converge weakly to that invariant distribution. ■

For any $\alpha_j, \beta_j, \omega_j \in (0, 1)$, we have established that the invariant distribution of firms sizes has thinner tails than the Pareto distribution with coefficient one. Moreover, we can order distributions in terms of the thinness of their tails, and can show that industries with higher physical capital shares have thinner tails. This will be useful below when we contrast the size distributions of firms in industries with different physical capital shares. We make these notions precise in the following definition and proposition.

Definition 5 *Let λ and ψ be probability measures on $[\underline{b}, \bar{b}]$. The probability measure λ has **thinner tails** than ψ if there exists \underline{x} and $\bar{x} \in [\underline{b}, \bar{b}]$ such that for all $\underline{b} \leq x \leq \underline{x}$, $\lambda([\underline{b}, x]) \leq \psi([\underline{b}, x])$, for all $\underline{x} \leq x \leq \bar{x}$, $\lambda([\underline{x}, x]) \geq \psi([\underline{x}, x])$, and for all $\bar{x} \leq x \leq \bar{b}$, $\lambda([\bar{x}, x]) \leq \psi([\bar{x}, x])$.*

In order to apply this definition, we need to standardize the support of the size distributions produced by our model. This is also necessary to contrast the implications of our model with the data where the size categories are the same for all industries. If we scale the productivity process A_{tj} by

$$\frac{1 - \omega_j (1 - \beta_j (1 - \alpha_j))}{\beta_j (1 - \alpha_j)}$$

the support of the firm size distribution is unchanged across industries and is equal to $[-\ln \bar{A}, -\ln \underline{A}]$. Under this scaling, we prove the following proposition.

Proposition 6 *For any $\alpha_j, \beta_j, \omega_j \in (0, 1)$, the invariant distribution of firm sizes has thinner tails than the Pareto distribution with coefficient one. Other things equal, if $\alpha_j > \alpha_k$, the invariant distribution of firms in sector j has thinner tails than the invariant distribution of firms in sector k .*

Proof. The first claim is immediate from the discussion above. To see the second, for each α denote the unique invariant probability measure of firm sizes (see Proposition 4) by $\lambda_\alpha : \mathcal{LN} \rightarrow [0, 1]$, where \mathcal{LN} denotes the Borel σ -algebra associated with LN , with associated transition function Q_α and operator T_α^* . Since λ_α is an invariant distribution

$$\begin{aligned} \lambda_\alpha([- \ln \bar{A}, \ln n]) &= (T_\alpha^* \lambda_\alpha)([- \ln \bar{A}, \ln n]) = \int Q_\alpha(z, [- \ln \bar{A}, \ln n]) \lambda_\alpha(dz) \\ &= \int \mu(A : g_\alpha(z, A) \in [- \ln \bar{A}, \ln n]) \lambda_\alpha(dz), \end{aligned}$$

where $g_\alpha(z, A)$ denotes the log firm size growth rate. We saw above that

$$\frac{dg_\alpha(z, A)}{d\alpha} < 0.$$

Then, for n small enough, we know that

$$\begin{aligned} \lambda_{\alpha_k}([- \ln \bar{A}, \ln n]) &= \int \mu(A : g_{\alpha_k}(z, A) \in [- \ln \bar{A}, \ln n]) \lambda_{\alpha_k}(dz), \\ &> \int \mu(A : g_{\alpha_j}(z, A) \in [- \ln \bar{A}, \ln n]) \lambda_{\alpha_k}(dz), \end{aligned}$$

and hence λ_{α_k} is not the invariant distribution α_k , and the operator $T_{\alpha_j}^*$ maps the λ_{α_k} into distributions with thinner left tails. The case for intermediate and high $\ln n$ are analogous. ■

In this section, we established that the process of accumulating industry specific human capital alone is sufficient to generate many observed properties of firm size dynamics and firm size distributions. In particular, mean reversion in the stock of industry specific human capital will cause small firms to grow faster than large firms and exit rates of firms to decline with size. Moreover we were also able to establish that the invariant distribution of firm sizes would have thinner tails than the Pareto distribution with coefficient one.

As a consequence of using the accumulation of industry specific human capital to explain scale dependence, our theory also predicts that the degree of scale dependence varies with the physical capital intensity of the industry. In Section 4 below we examine this implication in US data. Before turning to the data, the next section establishes that these implications are robust to a number of different modelling assumptions that were adopted above either for simplicity or expositional reasons.

3. 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, to the introduction of firm level heterogeneity, to alternative mechanisms for the accumulation of human capital such as learning by doing, 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 human capital stocks leads to mean reversion in firms sizes.

3.1 Firm Costs

The basic mechanism of our paper relies on mean reversion in the stock of industry specific human capital of production. Mean reversion in turn 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 the stock of specific human capital grew. In other words, reversion to the mean in the stock of specific factors *from above*, produces reversion to the mean in firm sizes *from below*.

Changes in the specification of the cost structure 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 equation (5). Now, however, assume that fixed costs depend on the average number of workers hired in industry j at time t , \bar{n}_{tj} . In particular, assume that fixed costs are given by $F_j \bar{n}_{tj}^{\xi_j}$. We have in mind institutional or organizational costs (for example dealing with unions or other industry organizations) that depend on the average size of firms in the industry. Individual firms do not take into account the effect of their hiring decisions on the fixed costs, so the problem of the firm is identical to the one presented above. The problem of the firm is to maximize profits

$$\max_{k_{tj}, h_{tj}, n_{tj}} \Pi \equiv \max_{k_{tj}, h_{tj}, n_{tj}} y_{tj} - r_{tj} k_{tj} - s_{tj} h_{tj} - w_{tj} n_{tj} - F_j \bar{n}_{tj}^{\xi_j},$$

where r_{tj}, s_{tj}, w_{tj} denote the corresponding factor prices. We assume that $0 \leq \xi_j < 1$ and so if $\xi_j = 0$ we have the same case studied above. Taking first order conditions and allowing for free entry and exit so that profits are zero implies

$$(1 - \gamma_j) y_{tj} = F_j \bar{n}_{tj}^{\xi_j}.$$

Now output changes with the average level of employment in the industry and, since in equilibrium all firms are identical, also with the employment level of the firm. Given this symmetry, equilibrium in factor markets implies that the size of the typical firm in the industry is given by

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

This equation is 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 factor supplies.⁷ Moreover, the direction of the change can differ: for $\xi_j < \gamma_j$, the behavior of employment is as before, declining with the industry physical and human capital stocks; for $\xi_j > \gamma_j$ this pattern is reversed and the size of firms

⁷Notice that because the fixed costs entail an external cost, the equilibrium will not be Pareto optimal. However, one can set up a pseudo-social planner problem that yields the same aggregate implications than the problem discussed in the Section 2 (see also Section 3.4).

depends positively on the stock of both types of capital but negatively with industry employment. 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 human capital stocks are large or small they revert to the mean. The example illustrates that the necessary property of firm sizes is that they respond monotonically to the stock of human capital in the industry. The direction of this response is not important: in the case where $\xi_j > \gamma_j$, reversion to the mean in the stock of specific factors *from above*, produces reversion to the mean in firm sizes *from above*. Mean reversion in the stock of human capital then leads to the same arguments and results we presented above.

3.2 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 human capital to the evolution of firm sizes. Clearly, 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, where we suppress time and industry subscripts. Suppose that after having decided to produce in a period (that is, after paying the fixed cost F) each firm $i \in [0, \mu]$ observes a firm specific productivity shock z_i . This shock is assumed to be i.i.d. over time, firms and industries within a sector. After observing this shock, the firm i can then hire labor n_i and industry- j -specific physical, k_i , and human capital, h_i , to produce output according to

$$y_i = z_i \left(k_i^\alpha \left[h_i^\beta n_i^{1-\beta} \right]^{1-\alpha} \right)^\gamma.$$

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 amounts of industry specific physical and human capital are fixed at K and H . 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

$$\int_0^\mu z_i \left(k_i^\alpha \left[h_i^\beta n_i^{1-\beta} \right]^{1-\alpha} \right)^\gamma di,$$

subject to

$$\int_0^\mu k_i di \leq K, \quad \int_0^\mu h_i di \leq H, \quad \int_0^\mu n_i di \leq N.$$

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

$$\mu \int_0^1 y_i \phi(di),$$

subject to

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

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

$$\frac{k_i}{k_j} = \frac{h_i}{h_j} = \frac{n_i}{n_j} = \left(\frac{z_i}{z_j} \right)^{\frac{1}{1-\gamma}},$$

and relative outputs

$$\frac{y_i}{y_j} = \left(\frac{z_i}{z_j} \right)^{\frac{1+\gamma}{1-\gamma}}.$$

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

$$\frac{k_i}{K} = \frac{h_i}{H} = \frac{n_i}{N} = \frac{z_i^{\frac{1}{1-\gamma}}}{\mu \int_0^1 z_i^{\frac{1}{1-\gamma}} \phi(di)}.$$

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 , and previously

accumulated physical and human capital K and H as

$$\int_0^\mu z_i \left(k_i^\alpha \left[h_i^\beta n_i^{1-\beta} \right]^{1-\alpha} \right)^\gamma di = \left(K^\alpha \left[H^\beta N^{1-\beta} \right]^{1-\alpha} \right)^\gamma \mu^{1-\gamma}$$

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 , as well as investment in both types of capital, are analogously determined.

Clearly, the addition of an i.i.d. 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* human capital.

3.3 Learning-by-Doing Externalities

In the model of Section 2, we assumed that human capital accumulation required some industry specific inputs. The dependence on industry-specific inputs was important for our model, as it allows human capital accumulation to vary with output in the industry, and is the primary source of mean reversion at the industry level.

In that model, the inputs to learning were purchased by consumers, and the resulting level of human capital was rented out by consumers, so that there was no externality. An alternative assumption that has similar effects is the assumption that human capital is accumulated from *learning-by-doing externalities* of the form

$$H_{t+1j} = A_{t+1j} H_{tj}^{\omega_j} Y_{tj}^{1-\omega_j},$$

which states that the higher is output in the industry, the higher is accumulation of human capital. Importantly, this involves no resource cost to the economy. Suppose also that production occurs according to

$$Y_{tj} + F_j \mu_{tj} = [K_{tj}^{\alpha_j} (H_{tj} N_{tj})^{1-\alpha_j}]^{\gamma_j} \mu_{tj}^{1-\gamma_j},$$

so human capital operates exactly like labor augmenting technological progress.

Although we can no longer use the social planners problem to solve for equilibrium allocations in this model, but we can use a pseudo-planner problem to solve it as we do in Subsection 3.4. Similar reasoning then produces an expression for the normalized rate of the growth of the firm of

$$\ln n_{t+1} - \ln n_t = n^C - \alpha_j (1 - \omega_j) \ln n_t - (1 - \alpha_j) \ln A_{t+1},$$

where n^C again denotes a constant specific to this formulation. If there is no learning by doing, or $\omega_j = 1$, there is no mean reversion in human capital stocks, and firm growth rates satisfy Gibrat's Law. As before, increases in the capital intensity of an industry increase the rate of mean reversion in firm sizes.

This extension emphasizes that it is not industry specific human capital *per se*, but rather the sensitivity of current production decisions to past output in the industry, that is important for our results on mean reversion. This is important in the light of recent research by Kambourov and Manovskii (2002) who argue that there is little evidence for industry-specific human capital in individual earnings data⁸. However, this evidence is consistent with industry-specific learning-by-doing externalities where individual workers do not appropriate the returns to industry-specific human capital.

3.4 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

⁸Our model makes no distinction between workers within an industry, and so cannot distinguish between industry-specific human capital and the occupation-specific human capital emphasized by Kambourov and Manovskii (2002).

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 a version of this theory where both margins react to factor accumulation. In particular, a theory that includes the three factors that we introduced in the model above. Now, physical and human capital are specific to an industry but mobile across varieties within that industry.

3.4.1 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) physical and human 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 $\sigma_j > 1$ to produce a composite industry good that is used for investment in human capital and as an input to production of a final good (in combination with the composite goods of other industries) that is consumed and invested in physical capital.

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

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

subject to

$$\begin{aligned} E_0 \left[\sum_{t=0}^{\infty} \sum_{j=1}^J \int_{0 \leq \varpi \leq \Omega_{tj}} p_{tj\varpi} D_{tj\varpi} d\varpi \right] &\leq E_0 \left[\sum_{t=0}^{\infty} \sum_{j=1}^J r_{tj} K_{tj} + s_{tj} H_{tj} + w_{tj} N_{tj} \right], \\ K_{t+1j} &= K_{tj}^{\lambda_j} X_t^{1-\lambda_j}, \quad H_{t+1j} = A_{t+1j} H_{tj}^{\omega_j} I_{tj}^{1-\omega_j} \\ Q_{tj} + I_{tj} &\equiv E_{tj} \leq \left\{ \int_{0 \leq \varpi \leq \Omega_{tj}} (D_{tj\varpi})^{\frac{\sigma_j-1}{\sigma_j}} d\varpi \right\}^{\frac{\sigma_j}{\sigma_j-1}}, \\ C_t + X_t &= \prod_{j=1}^J (Q_{tj})^{\theta_j}, \quad \sum_{j=1}^J N_{tj} \leq N_t. \end{aligned}$$

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

Q_{tj} is the amount of the final good in industry j used to produce consumption and physical capital investment in combination with the goods in other industries. 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

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

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.

3.4.2 Firms and industry equilibrium.—

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

$$y_{\varpi} = k_{\varpi}^{\alpha} [h_{\varpi}^{\beta} n_{\varpi}^{1-\beta}]^{1-\alpha},$$

We suppress the time and industry subscripts whenever this does not lead to confusion. The first stage of the problem of the firm is to minimize costs,

$$\begin{aligned} C(r, s, w, D_{\varpi}, F) &\equiv \min_{K_{tj}^{\varpi}, L_{tj}^{\varpi}} rK_{\varpi} + sH_{\varpi} + wN_{\varpi} \\ \text{s.t. } D_{\varpi} + F &= k_{\varpi}^{\alpha} [h_{\varpi}^{\beta} n_{\varpi}^{1-\beta}]^{1-\alpha}, \end{aligned}$$

where D_{ϖ} is the quantity demanded of the variety and F is a fixed cost of production. The cost function of the problem then becomes

$$C(r, s, w, D_{\varpi}, F_j) = \lambda (D_{\varpi} + F),$$

where

$$\lambda = \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{s}{\beta(1-\alpha)}\right)^{\beta(1-\alpha)} \left(\frac{w}{(1-\beta)(1-\alpha)}\right)^{(1-\beta)(1-\alpha)}.$$

Notice that average costs $C(r, s, w, D_{\varpi}, F_j) / D_{\varpi}$ are a decreasing function of D_{ϖ} .

The second stage of the firm problem is to maximize profits

$$\Pi(r, s, w, F) = \max_{p_\varpi} D_\varpi(p_\varpi) p_\varpi - C(r, s, w, D_\varpi(p_\varpi), F),$$

where $D_\varpi(p_\varpi)$ is derived from the consumers problem and stated above.

The first order conditions of the firm problem then imply that $p_\varpi = \lambda\sigma/(\sigma - 1)$. Hence in equilibrium the levels of production and profits by firms are given by

$$D_\varpi(p_\varpi) = \frac{E}{\Omega\lambda} \frac{\sigma - 1}{\sigma} \quad \text{and} \quad \Pi(r, s, w, F) = \frac{E}{\sigma\Omega} - F\lambda.$$

Zero profits then implies that the number of varieties (or firms since only one firm produces each variety) is given by $\Omega = E/(\sigma F\lambda)$ and so

$$D_\varpi(p_\varpi) = F(\sigma - 1).$$

The equilibrium conditions in factor markets are given by

$$K = \frac{E\alpha}{r}, \quad H = \frac{E\beta(1 - \alpha)}{s}, \quad N = E \frac{(1 - \beta)(1 - \alpha)}{w},$$

which implies that

$$\lambda = EK^{-\alpha} H^{-\beta(1-\alpha)} N^{-(1-\beta)(1-\alpha)} \quad \text{and so} \quad \Omega = \frac{K^\alpha H^{\beta(1-\alpha)} N^{(1-\beta)(1-\alpha)}}{\sigma_j F_j}.$$

Output in the industry is given by

$$Y = \Omega D_\varpi(p_\varpi) = \frac{\sigma - 1}{\sigma} K^\alpha H^{\beta(1-\alpha)} N^{(1-\beta)(1-\alpha)}.$$

Notice that this function is constant returns to scale, with TFP given by a function of the elasticity of substitution.

The size of firms in terms of employees is given by

$$n_\varpi = F\sigma \left(\frac{N}{K}\right)^\alpha \left(\frac{N}{H}\right)^{\beta(1-\alpha)},$$

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.

3.4.3 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

$$C_t + X_t \leq \prod_{j=1}^J \left(\frac{\sigma_j - 1}{\sigma_j} K_{tj}^\alpha H_{tj}^{\beta(1-\alpha)} N_{tj}^{(1-\beta)(1-\alpha)} - I_{tj} \right)^{\theta_j},$$

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 to produce investment in physical and human capital. 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 (with $\gamma = 1$) 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.

4. 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 human capital stocks, and therefore in firms sizes, increases with the degree of diminishing returns in human capital, or equivalently decreases with the degree of diminishing returns in physical capital. A very low physical capital share implies a high human capital share, hence a low degree of diminishing returns in human capital and, therefore, a low degree of reversion to the mean in firm sizes. As the physical capital share increases from zero

the degree of diminishing returns in human capital increases as does the reversion to the mean in firm sizes. 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 physical capital in technology. In this section we contrast this implication with the data.

4.1 Data

We have investigated the variation in scale dependence across sectors using data on growth rates and the distribution of firm sizes. We use two data-sets constructed especially for us by the US Census Bureau. The first is a data-set 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 data-set, 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 physical capital shares on the size distribution of firms. Variations in physical 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 is a better fit for our theory. The unique aspect of the longitudinal data-set 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 physical 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 which include human capital. We then construct physical capital shares as one minus the labor share. This implies that the physical capital shares we use include everything that is not classified as labor. There are two potential problems with the physical capital shares we compute. First, the physical 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 physical capital shares we use. Second, we are using the physical capital share in value added, but our theory is abstracting from the use of intermediate inputs. To address the former, we only consider industries with physical capital shares smaller than one half, although the results are similar if we consider all sectors. To address the latter, we also present results with physical capital shares adjusted for the share of value added and the share of materials purchased from the same industry.

4.2 Growth Rates

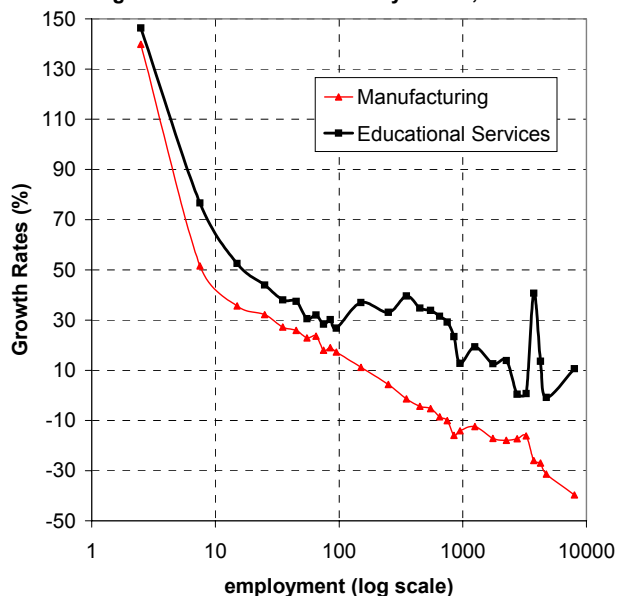
We begin by examining the growth rates of surviving firms. As a first step, consider an example with two sectors. Educational services is a very labor and human capital intensive sector with a physical capital share of 0.054, while manufacturing is much more physical capital intensive with a physical capital share of 0.397. If the theory is consistent with the data, given that manufacturing is more physical capital intensive, we should see growth rates of manufacturing firms decline faster with size than growth rates of firms in the educational sector (Proposition 1).

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. We examine next the implication of our theory that scale dependence in growth rates increases with physical capital shares (denoted by α_j) for all industries. We use data on the growth of firms, g_j , in a particular size category, x_j , and estimate the following regression:

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

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



This amounts to fitting an exponential trend where the parameter varies linearly with physical 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 $\tilde{\epsilon}$ should be negative and significant.

Table 1

	Exponential 1990-2000	Exponential 1990-2000 (adjusted)	Power 1990-2000	Power 1990-2000 (adjusted)
$\tilde{\epsilon}$	-0.0965	-0.1303	-0.2638	-0.3503
Standard error	0.0273	0.0345	0.0195	0.0250
P-value	0.0004	0.0002	0.0000	0.0000

The estimate of $\tilde{\epsilon}$ 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 physical capital shares: A doubling in the size of firms in manufacturing ($\alpha_j \approx 1/3$) decreases average growth by about 4% while in educational services ($\alpha_j \approx 0$) the growth rate is roughly the same.

As mentioned before, the physical 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 physical capital shares that are too large relative to the ones in gross output. Since our theory does not include materials, it is not designed to address this distinction. 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 physical capital share to obtain an adjusted physical capital share. If all intermediate inputs originated in the same sector, the original physical capital shares would equal the adjusted physical capital shares. If the rest of the materials used in production are homogeneous, the adjusted physical 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 human 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.⁹ The omission of intermediate inputs from 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 share of raw labor, the variance of productivity shocks, or the depreciation parameters, may account for some of the unexplained variation too.

The last ten years have witnessed a substantial decline in employment among large manufacturing establishments. A potential concern is that this may be driving the larger scale dependence observed in these sectors. To address this concern, we replicate the previous exercise for manufacturing and non-manufacturing sectors separately. The results presented in Table 2 show that this phenomenon is

⁹Adjusting the physical capital shares increases the number of sectors in our sample with physical capital shares below one-half from 44 to 52.

not driving the results in Table 1. The point estimates for both manufacturing and non-manufacturing are close to the ones for the whole economy. The estimates for non-manufacturing are highly significant as before. For manufacturing the estimates are less precise reflecting the smaller variation in physical capital shares among these sectors. This was precisely our original justification for using all sectors in the economy. Table 2 presents results only for exponential trends, the result for power trends are similar.

Table 2

	Exponential 1990-2000 Manufacturing (adjusted)		Exponential 1990-2000 Non-Manufacturing (adjusted)	
$\tilde{\epsilon}$	-0.1086	-0.2159	-0.0953	-0.1282
Standard error	0.4944	0.6624	0.0274	0.0346
P-value	0.8262	0.7446	0.0005	0.0002

Our estimation of \tilde{b} and $\tilde{\epsilon}$ assumes that both β_j and ω_j are constant across industries (call these values β and ω respectively). We can then use the estimates presented in Table 1 and 2, together with the estimates of \tilde{b} and Equation (13) to infer values for β and ω . The estimates of \tilde{b} for the exercises in the first two columns of Table 1 are: $-.23$ (s.e. $.01$), $-.24$ (s.e. $.01$). This values imply a share of specific human capital in labor services given by $\beta = .29$ and $.35$, and a share of investments in human capital production given by $1 - \omega = .33$ and $.35$, respectively. That is, the model, and the estimation above, imply that the share of labor services related to specific human capital is around one third. Thus, as we have argued in this paper, the share of specific human capital consistent with the scale dependence in firm dynamics is very significant. Other forms of human capital that are not industry specific, and therefore are associated with individuals and not with an industry are, of course, not included in this share. If some of the industry specific human capital is accumulated through learning by doing, and therefore not appropriated by individuals as in Section

3.3., these numbers will be biased down. The inferred share of investments in human capital production $(1 - \omega)$ is certainly reasonable given that we are using 10 year growth rates. We replicated the same exercise for the manufacturing and service sectors. The estimates of \tilde{b} for the exercises in the first and third column of Table 2 are: $-.20$ (s.e. $.15$) and $-.23$ (s.e. $.01$). This yields $\beta = .35$ and $1 - \omega = .31$ for manufacturing, as well as $\beta = .29$ and $1 - \omega = .33$ for services. Industry specific human capital is, therefore, more important in the service sector than in manufacturing and it depreciates more slowly, although the differences are small and insignificant in light of the imprecision of our estimates for the manufacturing sector. This is reassuring given our assumption that both β and ω do not vary significantly across industries.

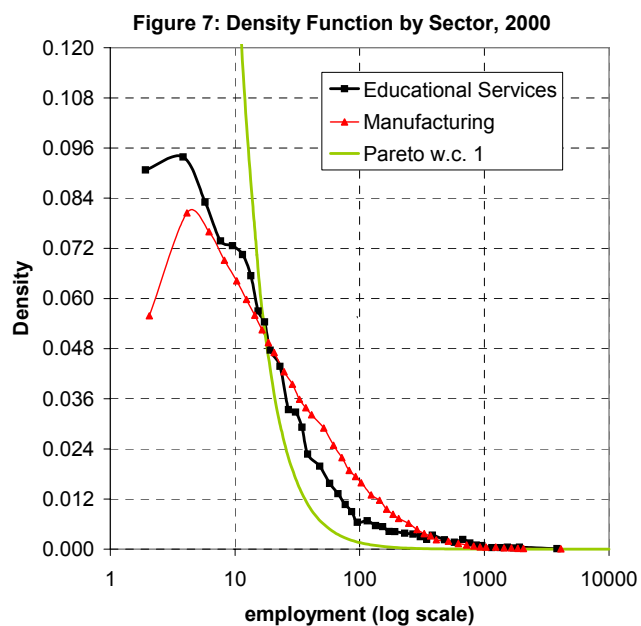
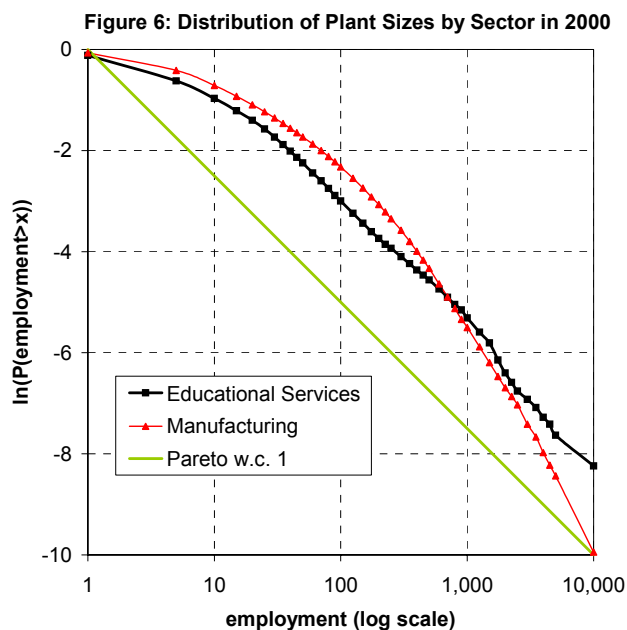
4.3 Size Distribution of Firms

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 physical capital shares (Proposition 6). A first look at the data is presented in Figures 6 where we plot $\ln P_j$ and x_j for educational services and manufacturing.

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 as long as $\beta_j, \omega_j, (1 - \alpha_j) > 0$. 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 physical 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. To examine this, we estimate the following regression

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

where \hat{a}_j and \hat{b}_j are industry specific coefficients. This amounts to constraining the quadratic term to vary linearly with the physical capital share. The model now predicts that \hat{e} should be negative and significant. The results are presented in Table 3.

The estimate of \hat{e} for 1990, in the first column of Table 3, is negative and strongly significant. We also estimated the same regression using NAICS three digit sectors in 2000. The physical 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 3 show that the estimate of \hat{e} is smaller in absolute value but still negative and strongly significant. The results with adjusted physical capital shares are presented in the third column of Table 3, which further confirms the empirical significance of the mechanism in our theory. As for growth rates the results are similar when we use the sub-samples of manufacturing and non-manufacturing sectors.

Table 3

	1990	2000	1990 (adjusted)
\hat{e}	-0.0776	-0.0352	-0.0580
Standard error	0.0069	0.0019	0.0074
P-value	0.0000	0.0000	0.0000

4.4 Exit Rates

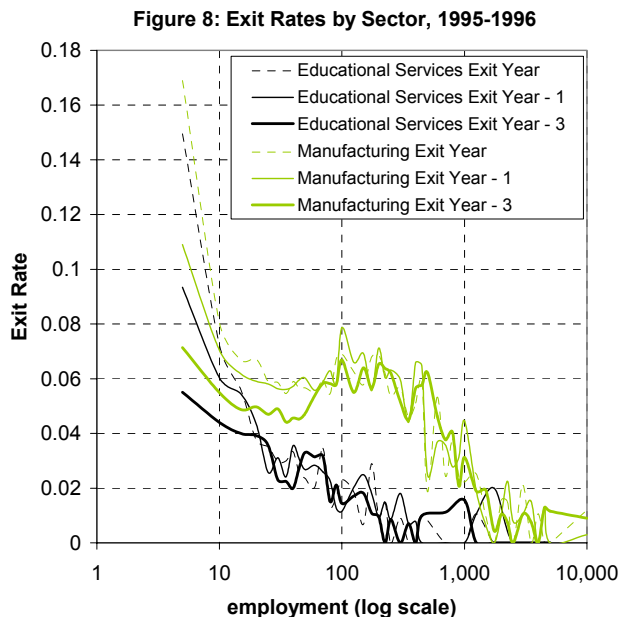
Our mechanism, which emphasizes mean reversion in stocks of specific factors, when combined with particular assumptions on preferences, also implies that exit rates should decline with firm size. Furthermore, the rate of decline should vary with physical capital shares. Figure 8 illustrates this using BITS data for US manufacturing and educational services in 1995-1996. The dashed 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 dark line by 1992 size class. The number of firm deaths is divided by the number of surviving firms to compute exit rates.

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 exponential trend in manufacturing is steeper than in educational services, although the difference is small given the large variance (a rate of decline of -0.0379 for man-

ufacturing and -0.0364 for educational services).¹⁰ The results hold more strongly across all sectors in the economy, especially if we focus on the size distribution of exiting firms three years before they exit. The reason we believe this is the best test for our theory is that it reduces exit produced purely by selection. We run the following regression

$$\ln(1 + ER_j) = \check{\alpha}_j + \check{\beta} \ln x_j + \check{\alpha}_j \ln x_j + \check{\varepsilon}_{tj},$$

which amounts to estimating an exponential relationship between exit rates and sizes three years before exit.



The results are presented in Table 4. The first two columns present the weighted least squares results with weights given by the total number of firms in the representative size category, for both unadjusted and adjusted physical capital shares. The last two columns present results from the same exercise using a power function instead of an exponential. The results are consistent with our theory: All of the estimates are negative and significant. The results are also economically significant: A doubling of

¹⁰Orr (1974), Gorecki (1976), Hause and Du Rietz (1984) and MacDonald (1986) found that firm exit rates were negatively related to measures of physical 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 physical capital intensive sectors are larger on average.

firm size decreases exit rates by around 0.3% in manufacturing while exit rates do not decline with size in educational services.

Table 4

1995 Exit Rates by Size in 1992				
	Exponential		Power	
	(adjusted)		(adjusted)	
$\check{\epsilon}$	-0.0092	-0.0164	-0.0136	-0.0233
Standard error	0.0009	0.0011	0.0013	0.0016
P-value	0.0000	0.0000	0.0000	0.0000

5. CONCLUSION

In this paper we have constructed a theory that is consistent with some well known facts on scale dependence in firm dynamics and firm size distributions. The mechanism emphasizes the role of the accumulation of industry specific human capital. 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 industry specific human capital in the theory led us to focus on cross sectoral differences in the importance of human, and therefore physical, capital in production, and in particular physical capital intensity. Increases in the importance of industry specific physical capital lead to an increase in the degree of diminishing returns in human capital, and hence more scale dependence in growth, exit rates, and firm size distributions. 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 the 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 is frictionless: a strong assumption especially if we look at firm sizes 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.

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 may be 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. Our results are, in general, not sensitive to government policies that affect firms independently of their size. Scale dependent policies may affect some of our implications and Restuccia and Rogerson (2004) argue that scale dependent policies may have large effects on efficiency. 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 human capital, our theory also makes predictions 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. 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] Cooley, T. and V. Quadini (2001) "Financial Markets and Firm Dynamics", *American Economic Review*, 91(5):1286-1310.
- [8] 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.
- [9] 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.
- [10] Ericson, R. and A. Pakes (1995). "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1): 53-82.
- [11] Evans, D. S. (1987a). "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy*, 95(4): 657-674.
- [12] 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.
- [13] Garicano, L. and E. Rossi-Hansberg (2004). "Inequality and the Organization of Knowledge." *American Economic Review P&P*, forthcoming.
- [14] Geroski, P. (1995). "What do we Know About Entry?" *International Journal of Industrial Organization*, 13, 421-440.

- [15] Geroski, P. (1983). "The Empirical Analysis of Entry: A Survey." University of Southampton, mimeo.
- [16] Geroski, P. and J. Schwalbach (1991). *Entry and Market Contestability*. Oxford, Blackwell.
- [17] Gibrat, R. (1931). *Les inégalités économiques*, Libraire du Recueil Sirey.
- [18] Gorecki, P. K. (1976). "The Determinants of Entry by Domestic and Foreign Enterprises in Canadian Manufacturing." *Review of Economics and Statistics*, 58: 485-488.
- [19] Griliches, Z. and H. Regev (1995). "Firm productivity in Israeli industry 1979-1988." *Journal of Econometrics*, 65: 175-203.
- [20] 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.
- [21] Hause, J. C. and G. Du Rietz (1984). "Entry, Industry Growth, and the Microdynamics of Industry Supply." *Journal of Political Economy*, 92: 733-757.
- [22] Hopenhayn, H. A. (1992). "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(2): 1127-1150.
- [23] Jones, C.I. (2004). "The Shape of Production Functions and the Direction of Technical Change," NBER working paper no. 10547.
- [24] Jovanovic, B. (1982). "Selection and the Evolution of Industry." *Econometrica*, 50(3): 649-670.
- [25] Kambourov, G. and I. Manovskii (2002). "Occupational Specificity of Human Capita." Unpublished paper, University of Pennsylvania.
- [26] Klette T. and S. Kortum, 2004, "Innovating Firms and Aggregate Innovation", forthcoming *Journal of Political Economy*, 112(5).
- [27] Luttmer, E. (2004). "The Size Distribution of Firms in an Economy with Fixed and Entry Costs." Working Paper 633, Federal Reserve Bank of Minneapolis.
- [28] MacDonald, J. M. (1986). "Entry and Exit on the Competitive Fringe." *Southern Economic Journal*, 52: 640-658.
- [29] Mansfield, E. (1962). "Entry, Gibrat's Law, Innovation, and the Growth of Firms." *American Economic Review*, 52(5): 1023-51.
- [30] Orr, D. (1974). "The Determinants of Entry: A Study of the Canadian Manufacturing Industries." *Review of Economics and Statistics*, 56: 58-66.

- [31] Restuccia, D. and R. Rogerson (2004). "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." Unpublished paper, Arizona State University.
- [32] Rossi-Hansberg, E. and M. L. J. Wright (2004). "Urban Structure and Growth." Unpublished paper, Stanford University.
- [33] Scherer, F. M. (1980). *Industrial Market Structure and Economic Performance*. Boston, Houghton Mifflin.
- [34] 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.
- [35] Sutton, J. (1997). "Gibrat's Legacy." *Journal of Economic Literature*, 35(1): 40-59.
- [36] 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.