

Establishment Size Dynamics in the Aggregate Economy

By ESTEBAN ROSSI-HANSBERG AND MARK L. J. WRIGHT*

This paper presents a theory of establishment size dynamics based on the accumulation of industry-specific human capital that simultaneously rationalizes the economy-wide facts on establishment growth rates, exit rates, and size distributions. The theory predicts that establishment growth and net exit rates should decline faster with size, and that the establishment size distribution should have thinner tails, in sectors that use specific human capital less intensively. We establish that there is substantial cross-sector heterogeneity in US establishment size dynamics and distributions, which is well explained by relative factor intensities. (JEL L11, L16, L25).

Establishment size dynamics are *scale dependent*: small establishments grow faster than large establishments conditional on survival, and net exit rates decline with size.¹ Scale dependence in growth and net exit rates is also systematically reflected in the size distribution of establishments. In this paper, we propose an explanation for this scale dependence which relies on the response of production decisions to the accumulation and allocation of industry-specific human capital. Our 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 dataset to document these facts for the US economy. We find that, as predicted by our theory, US sectors with larger physical capital shares exhibit significantly more scale dependence in establishment size dynamics and distributions.

Our basic approach is simple and starts by noting that all of the facts above are manifestations of *mean reversion* in the economy; indeed, the fact that, conditional on survival, small establishments grow faster than large establishments is an explicit statement of mean reversion. Moreover, mean reversion in factor accumulation is a general result in macroeconomic models. We focus on the accumulation of industry-specific human capital, because this type of capital is, as a result of on-the-job training and learning-by-doing, more closely tied to production conditions in an industry than is any other factor. In our theory, under standard conditions, an abundance of human capital leads to low rates of return and slower accumulation. Conversely, a relatively small stock of human capital leads to high rates of return and faster accumulation. 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 establishment sizes respond monotonically to fluctuations in factor prices, which are driven by the stock of human capital, mean reversion in these stocks leads to mean reversion in establishment sizes. Hence, conditional on survival, small establishments grow faster than large establishments.

* Rossi-Hansberg: Department of Economics, Princeton University, Princeton, NJ, 08544 (e-mail: erossi@princeton.edu); Wright: Department of Economics, University of California, Los Angeles, Los Angeles, CA, 90095 (e-mail: mlwright@econ.ucla.edu). We thank the editor, two referees, Liran Einav, Gene Grossman, Bob Hall, Brian Headd, Tom Holmes, Boyan Jovanovic, Gueorgui Kambourov, Pete Klenow, Narayana Kocherlakota, Per Krusell, Erzo Luttmer, Steve Redding, Kathy Rolfé, and numerous seminar participants for helpful comments; Tim Bresnahan and CEEG for financial support; Trey Cole of the US Census Bureau for help in constructing the data; and Adam Cagliarini and Raphael Godefroy for outstanding research assistance. We acknowledge financial support from the National Science Foundation Grant SES-0453125.

¹ This fact was most forcefully demonstrated by Edwin Mansfield (1962) in his study of firms in the steel, petroleum, tire, and automobile industries. More recent work by David S. Evans (1987a, b) and Bronwyn H. Hall (1987) using data on firms, and by Timothy Dunne, Mark J. Roberts, and Larry Samuelson (1989a, b) using data on manufacturing plants, has confirmed this finding. See also the surveys by Frederic M. Scherer (1980), Paul A. Geroski (1995), John Sutton (1997), and Richard E. Caves (1998), which document the robustness of these results across time, industries, and countries.

The same process generates net exit rates that decline with size. To see this, note that given the level of employment in an industry, increases in average establishment size imply that some establishments have exited, while decreases imply that some establishments have entered. This logic is preserved as long as employment in the industry does not increase with a positive productivity shock by more than average establishment sizes, which will be the case as long as the elasticity of substitution between goods in consumption is not too large. If so, since small establishments grow faster than large establishments, the net exit rate is largest for small establishments, and we have scale dependence in net exit rates. We can then combine the implications of the model for growth and net exit to show that in the long run the distribution of establishment sizes in a sector converges to an invariant distribution that displays scale dependence in the sense that it has thinner tails than the Pareto distribution with coefficient one.

Our emphasis on the accumulation and allocation of specific human capital implies that establishment growth and exit rates should decline faster with size in sectors that use human capital less intensively. This is intuitive: the less intensively human capital is used, the faster diminishing returns to scale set in and the faster the rate of mean reversion. In turn, this implies that the tails of the size distribution of establishments should be thinner the smaller the human capital share. Hence, the degree of mean reversion decreases with human capital intensity, just as in the neoclassical growth model the speed of convergence decreases with the physical capital share. We show that the process of entry and exit of establishments ensures that industry production will display constant returns to scale, 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 establishment sizes.

We assess the actual relationship between capital shares and establishment scale dependence in the United States using a new dataset commissioned from the US Census Bureau on establishment growth and net exit rates, as well as establishment size distributions, for very fine size categories and two-digit Standard

Industrial Classification (SIC) (or three-digit North American Industry Classification System (NAICS)) sectors. Using these data, we first test the implication for growth rates and show that, as predicted by the theory, scale dependence in establishment growth rates is positively and significantly related to physical capital shares. We then show that this relationship is reflected in net exit rates and in significant differences in the size distribution of establishments across sectors. The differences are economically large. For example, a doubling of the size of an establishment results in a decline in growth rates of more than half a percentage point per year in the physical capital-intensive manufacturing sector, but has little effect in the labor-intensive educational services sector. Likewise, in order to make the size distribution of establishments in the manufacturing sector conform to the size distribution of establishments in the educational services sector, we would need to move roughly 3 million employees (about 20 percent of total manufacturing employment) out of medium-sized manufacturing establishments (between 50 and 1,000 employees), and reallocate 2 million of them to very large establishments and 1 million to very small establishments. We believe that this is the first study to use such detailed establishment size data for the entire nonfarm private sector. The broad, fine coverage allows us to uncover the novel empirical regularities predicted by our theory.²

Most recent theoretical attempts to explain the size dynamics and distribution of establishments generate scale dependence via selection mechanisms: unsuccessful establishments decline and exit. Hugo A. Hopenhayn (1992), Richard Ericson and Ariel Pakes (1995), and Erzo G. J. Luttmer (forthcoming) all model selection as the

² Relatively little work has examined cross-industry differences in establishment sizes. In terms of firm/plant growth rates, D. B. Audretsch et. al. (2004) find that Gibrat's law is a better approximation for the Dutch services sector than for the manufacturing sector. In terms of entry and exit, Geroski (1983) finds that gross entry and exit rates of firms are positively correlated across industries, while Geroski and Joachim Schwalbach (1991) find that turnover rankings are common across countries. Dale Orr (1974), Paul K. Gorecki (1976), John C. Hause and Gunnar Du Rietz (1984), and James M. MacDonald (1986) all find that firm/plant exit rates are negatively related to measures of physical capital intensity by industry.

result of sequences of bad productivity shocks, while in the model of Boyan Jovanovic (1982) selection occurs when establishments learn about their fixed productivity, and in that of Tor Jakob Klette and Samuel Kortum (2004) as establishments adjust product lines in response to their own and competitors' investments in research and development. In contrast, while acknowledging that a selection mechanism is important for small establishments, we argue that it is less relevant in explaining the scale dependence observed for medium-sized and large establishments, and we abstract from it in our theory.

Another mechanism that generates scale dependence in establishment dynamics is the presence of inefficiencies in financial markets, as in the models of Thomas F. Cooley and Vincenzo Quadrini (2001), Rui Albuquerque and Hopenhayn (2002), Luis M. B. Cabral and José Mata (2003), and Gian Luca Clementi and Hopenhayn (2006). Other models, for example those of Robert E. Lucas, Jr. (1978) and Luis Garicano and Rossi-Hansberg (2004), produce a size distribution for establishments that inherit the properties of the distribution of managerial ability in the population. In contrast, our approach *endogenously* produces the size dynamics and distribution of establishments as the result of the *efficient* accumulation and allocation of factors of production.

Our theory is not the first to successfully produce scale dependence in establishment growth rates, net exit rates, and the size distribution observed for all establishments in the US economy. However, many of the other theories have very different implications for welfare and government policy. Consequently, we need to find new dimensions of the data which we can use to discriminate among these theories. Here, we propose such a dimension: the variation in scale dependence across sectors. We derive the empirical predictions of our theory and show that, consistent with the theory, scale dependence in growth rates, net exit rates, and the size distribution increases with physical capital shares. None of the other theories has developed this prediction. Paraphrasing Jovanovic (1982), many of the mechanisms in the literature undoubtedly contribute toward an explanation of establishment 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 I describes our theory and derives its key empirical predictions. Section II describes our data and shows that establishment growth and net exit rates, as well as the establishment size distribution, vary with physical capital shares in precisely the way predicted by our theory. Section III concludes. A number of extensions, designed to show the robustness of our mechanism and its predictions to changes in the institutional environment, are presented in Appendix A, along with a discussion of the link between our theory and the empirical work on specific human capital by Gueorgui Kambourov and Iouri Manovskii (2005). Appendix B contains proofs of the propositions.

I. The Model

We present a stochastic dynamic general equilibrium model in which establishments are perfectly competitive. Labor is mobile across industries, while both physical and human capital are specific to each industry. At an establishment, fixed costs plus increasing marginal costs of production imply a U-shaped average cost curve, while free entry and exit ensures that all establishments operate at the bottom of their average cost curves. Since our focus is on the accumulation and allocation of factors of production, the demand side of the model is kept as simple as possible by assuming logarithmic preferences. Combined with Cobb-Douglas production functions and log-linear depreciation, this ensures that we can solve the entire model in closed form.

A. Households

The economy is populated by a unit measure of identical small households. At the beginning of time, each 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

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

where δ is the discount factor of the household and E_0 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 individual valuations of all of its members.

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

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

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 the economy has S sectors and that each sector has J_s industries, where $s = 1, \dots, S$. Each industry produces a single distinct good, so that the number of goods being produced in this economy is $J = \sum_{s=1}^S J_s$. 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 identical. Each industry within a sector is assumed to have 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 useful since then each industry within a sector evolves separately, according to a process governed by the same parameters, the invariant distribution of establishment sizes within each sector can be characterized. In thinking about the data, we will define our sectors to be roughly comparable to the list of three-digit NAICS classifications, while our industries will map into NAICS industries at a much finer level of disaggregation.

In each time 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 the amount of

time worked in industry j in period t , N_{tj} is constrained by

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

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

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

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:

$$(5) \quad H_{t+1j} = A_{t+1j} H_{tj}^{\omega_j} I_{tj}^{1-\omega_j}.$$

Here, A_{t+1j} is an industry-specific productivity shock that is assumed to be independent and identically distributed (i.i.d) with compact support $[\underline{A}_j, \bar{A}_j]$ and is designed to capture the random accumulation of knowledge within an industry. I_{tj} denotes investment in human capital accumulation. This industry-specific productivity shock is the only source of randomness in our model and follows processes that are common across sectors.³ We assume that I_{tj} is denominated in terms of the output of the particular industry, in order to capture the idea that industry-specific learning requires some industry-specific inputs. The resource constraint for output of industry j , Y_{tj} is therefore $Q_{tj} + I_{tj} = Y_{tj}$.

In our framework there are no externalities: human capital investments are paid by households, and they rent the new human capital

³ We could have added industry-specific shocks to total factor productivity (TFP), instead of shocks to the human capital accumulation equation. This would not change any of our substantive results, but would come at the cost of some substantially more complicated algebra.

for use in production. In Appendix A, we will extend the model to allow for learning-by-doing externalities and show that this extension has similar properties. Moreover, with learning-by-doing externalities, households do not appropriate the rewards to industry-specific learning, which is consistent with the empirical evidence on industry-specific human capital, as seen, for example, in the work of Kambourov and Manovskii (2005). The assumption that human capital accumulation responds to industry-specific production levels is essential for our results because it is the primary source of industry-specific mean reversion.

Finally, as discussed 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 K_{0j} and H_{0j} .

B. Establishments

Production within each industry takes place in production units that we call *establishments*. For simplicity, we initially abstract from establishment-specific heterogeneity and assume that each establishment in industry j in period t has access to the same production technology (we will relax this assumption in Appendix A). To produce in any period, an establishment must pay a fixed cost F_j that period. Once that cost has been paid, the establishment hires industry- j -specific physical capital k_{ij} , in combination with an industry- j -specific labor input that is, in turn, produced by combining raw labor n_{ij} with industry- j -specific human capital, h_{ij} , and produces according to

$$(6) \quad y_{ij} = k_{ij}^{\alpha_j} (h_{ij}^{\beta_j} n_{ij}^{1-\beta_j})^{1-\alpha_j} \gamma_j.$$

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 pins down the size of the establishment. 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. Production parameters 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 depends on the denomination of the fixed cost, so to begin we assume it is denominated in the units of the establishment's output. This has the expositional advantage of pinning down the scale of production of the establishment (measured in terms of output), so that we can easily analyze the effects of changes in factor prices on the size of the establishment (measured in terms of the number of employees); we return to this assumption in Appendix A.

C. Capital Accumulation and Labor Allocation

To complete the characterization of the evolution of establishment sizes in this economy, all we need to do is characterize the evolution of productivity and factors in equilibrium. If we allow for a noninteger number of establishments, μ_{ij} , this economy satisfies all the assumptions of the welfare theorems. Because we are primarily interested in allocations, not prices, we proceed by solving the *social planning problem* for this economy: choose state-contingent sequences $\{C_{ij}, X_{ij}, I_{ij}, N_{ij}, \mu_{ij}, H_{ij}, K_{ij}\}_{t=0, j=1}^{\infty, J}$ so as to maximize household welfare

$$(7) \quad (1 - \delta)E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \ln \left(\frac{C_t}{N_t} \right) \right]$$

subject to, for all periods and states, the resource constraint on the final good

$$(8) \quad C_t + \sum_{j=1}^J X_{ij} = B \prod_{j=1}^J (Y_{ij} - I_{ij})^{\theta_j},$$

and the resource constraint on each intermediate good

$$(9) \quad Y_{ij} = \left[K_{ij}^{\alpha_j} (H_{ij}^{\beta_j} N_{ij}^{1-\beta_j})^{1-\alpha_j} \right]^{\gamma_j} \mu_{ij}^{1-\gamma_j} - F_j \mu_{ij},$$

for all industries; the accumulation equations for each industry-specific factor given by (4) and (5), and the constraint on labor allocation (3).

Inspection of this problem reveals that the choice of the number of establishments is entirely static: μ_{ij} appears only in the resource constraint for industry j in period t , (9). This implies that we can solve for the optimal number of establishments before solving for the dynamics of the economy. The first-order condition with respect to μ_{ij} is given by $F_j = (1 - \gamma_j) y_{ij}$, which leads to

an equilibrium establishment size (and a number of establishments, μ_{ij}) that depends on the amount of factors in the industry according to

$$(10) \quad n_{ij} = \frac{N_{ij}}{\mu_{ij}} = \left[\frac{F_j}{1 - \gamma_j} \right]^{\frac{1}{\gamma_j}} \left(\frac{N_{ij}}{K_{ij}} \right)^{\alpha_j} \left(\frac{N_{ij}}{H_{ij}} \right)^{\beta_j(1-\alpha_j)}.$$

This equation shows that, if the stock of specific factors is large relative to the amount of labor employed in the industry (which corresponds to a time of relatively cheap specific factor prices), then establishment size measured in terms of the number of employees will be small as establishments substitute toward the cheaper factors. Similarly, mean reversion in the stock of relative specific factor stocks will drive mean reversion in establishment sizes. Importantly, the qualitative nature of the relationship between factor stocks and establishment size can be reversed without changing the result that mean reversion in these stocks produces mean reversion in establishments size. We show below that the incentive to accumulate specific factors produces precisely the required mean reversion in the general equilibrium of our model.

Substituting for the optimal number of establishments μ_{ij} in the resource constraint gives

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

This is our first main result: by varying the number of establishments, each of which produces at the bottom of its average cost curve, we see that the industry behaves as if it has constant returns to scale. Hence, at the industry level (but not at the establishment level), increases in physical capital shares are related to decreases in human capital shares. The result is an entirely standard log-linear, multisector 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 (as in, for example, Rossi-Hansberg and Wright (2007)) 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, 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.

D. Establishment Growth, Net Exit, and the Size Distribution

We can now characterize the evolution of establishment sizes in the economy. Taking natural logarithms and differences of the expression for establishment size (10), we find that the growth rate of an establishment in industry j that survives from one period to the next is given by

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

Substituting for the evolution of human capital, we get that

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

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

⁴ In a related paper, Charles I. Jones (2005) shows how a Pareto size distribution of ideas leads to an aggregate Cobb-Douglas production function.

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 on 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_{ij} : if industry output is high, then human capital accumulation proceeds, on average, at a faster pace.

Before turning to a discussion of scale *dependence* in growth rates, we examine the conditions under which we get scale *independence*; in other words, the conditions under which we get Gibrat's law. Suppose we eliminate human capital as a factor of production by either reducing the importance of labor as a whole, $(1 - \alpha_j) \rightarrow 0$, or reducing the importance of human capital in producing labor services, $\beta_j \rightarrow 0$. Without human capital, establishments grow at a deterministic rate independent of scale. This is because the only source of industry-specific randomness comes from shocks to the accumulation of human capital.⁵ Alternatively, 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 establishments now becomes

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

⁵ One 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.

⁶ If $\omega_j = 1$, 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 $1 - \omega_j$ as arbitrarily small but positive.

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

To see how the growth rates of surviving establishments depend on establishment size in general, assume as before that population growth is zero and the aggregate economy is in steady state, so that physical capital is constant in all industries. Then, using equation (10) we can write the growth rate of the establishment, after substituting for I_{ij} , as

$$\begin{aligned} (11) \quad \ln n_{t+1j} - \ln n_{tj} &= n_j^C \\ &\quad - (1 - \omega_j)(1 - \beta_j + \alpha_j \beta_j) \ln n_{tj} \\ &\quad - \beta_j(1 - \alpha_j) \ln A_{t+1j}, \end{aligned}$$

where n_j^C is a constant term that depends on the physical capital stock. In steady state, the theory implies that the natural logarithm of establishment size is an AR(1) process with an autoregressive coefficient given by $1 - (1 - \omega_j) \times (1 - \beta_j + \alpha_j \beta_j) \leq 1$.

We summarize the results of this discussion in the following proposition. There, we emphasize the effect of changes in physical capital intensity, an observable parameter which we will focus on in our empirical analysis.

PROPOSITION 1: *Growth rates of surviving establishments are weakly decreasing in size. The larger the physical capital share, the faster growth rates decline with size. The growth rate of surviving establishments is independent of size only if either human capital is not a factor of production (in the limit when β_j or $(1 - \alpha_j)$ equal zero), 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 is constant over time. Combined with the result of Proposition 1, this has strong implications for net exit rates: net exit is positive whenever establishment sizes grow on average and negative when they decline. Moreover, Proposition 1 implies that the larger the physical capital share, the faster the net exit rate decreases with establishment size. In a more general model in which the labor allocation varies across industries in equilibrium, these

results continue 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 establishment sizes. We formalize these arguments as a corollary.

COROLLARY 2: *Establishment net exit rates are weakly decreasing in size. The larger the physical capital share, the faster net exit rates decline with size. The net exit rate of establishments is independent of size only if either human capital is not a factor of production (in the limit, when β_j or $(1 - \alpha_j)$ equal zero), or human capital evolves exogenously (in the limit, as ω_j approaches one).*

These implications for the relationship among physical capital shares, surviving establishment growth rates, and net exit rates can be tested directly using longitudinal data. In combination with the assumption that the distribution of establishment sizes has converged to its long-run distribution, these implications can also be tested with data on the size distribution of establishments. The next four propositions characterize the implications of our model for the invariant distribution of establishment sizes for different assumptions about parameter values, the distribution of human capital shocks, and the presence of a lower bound on establishment sizes. Obviously, as our economy is growing, we will need to normalize the sizes of establishments (or detrend their growth rates) in order to ensure the existence of an invariant distribution. The results in Propositions 3, 4, and 5 should be interpreted as applying to the appropriate normalized establishment sizes.

To begin, we examine the conditions under which our model is capable of reproducing the commonly used benchmark of Zipf's law: the size distribution is Pareto with coefficient one. A number of authors, including most notably Xavier Gabaix (1999), have shown that if Gibrat's law characterizes the growth rate of a finite number of establishments, and one imposes a lower bound on establishment sizes that converges to zero, then the invariant distribution converges to Zipf's law. In the present framework, the entry and exit of establishments means that these results do not directly apply. However, in a related paper, Rossi-Hansberg

and Wright (2007) show that scale-independent growth for a finite number of industries, combined with this form of entry and exit and a lower bound for establishment sizes that converges to zero, is sufficient to generate an invariant distribution that satisfies Zipf's law. An analogous result holds for the current framework, for the same limiting parameter values that produced Gibrat's law for establishment growth above.

PROPOSITION 3 (Zipf's Law): *If either human capital is not a factor of production (in the limit when β_j or $(1 - \alpha_j)$ equal zero), or human capital evolves exogenously (in the limit, as ω_j approaches one), and establishment sizes are bounded below by ℓ , the invariant size distribution of establishments converges to a Pareto distribution with shape coefficient one as ℓ converges to zero.*

Away from these limiting parameter values, so that there is mean reversion in conditional establishment growth rates, we characterize the properties of the invariant distribution for two cases. First, we examine a case in which productivity shocks are unbounded and are drawn from a lognormal distribution. In this special case, the invariant distribution of establishment sizes can be derived in closed form, and we can study the way its variance changes with physical capital shares. Second, we characterize the invariant distribution of establishment sizes for arbitrary productivity shock processes with bounded support. Here, we study how the amount of dispersion in the establishment size distribution—measured by the amount of mass in the tails of the distribution—varies with the capital share. This alternate measure of dispersion has the advantage that it is less sensitive to the sizes of the very largest establishments, which is especially important for combinations of parameters that are close to the limiting cases studied in Proposition 3, where the long-run variance of establishment sizes diverges.

To begin with the first case, assume that the logarithm of the productivity shock A_{jt} is distributed normally with mean M_{A_j} and variance $S_{A_j}^2$. Given the AR(1) form of the equation governing the evolution of surviving establishments, (11), it is straightforward to see that the invariant distribution of *representative* establishment sizes in a sector, in logarithms, will

be normal with mean $M_j = \beta_j (1 - \alpha_j) M_A$ and variance

$$(12) \quad \text{Var}(\ln n_j) \equiv S_j^2 = \frac{[\beta_j(1 - \alpha_j)]^2 S_{A_j}^2}{1 - [1 - (1 - \omega_j)(1 - \beta_j(1 - \alpha_j))]^2}.$$

To obtain the size distribution of establishments and its variance, we must also account for the process of entry and exit or, more specifically, adjust for the fact that an industry in this sector has precisely $\mu_j = N_j/n_j$ establishments of size n_j . The following proposition establishes that the actual size distribution of establishments (as opposed to the distribution of representative establishment sizes) also turns out to be lognormal.

PROPOSITION 4 (Lognormality): *If the productivity shock A_{ij} is distributed lognormally with mean M_{A_j} and variance $S_{A_j}^2$, then the long-run size distribution of establishments is lognormal with mean and variance, respectively, given by*

$$e^{M_j - (S_j^2/2)} \quad \text{and} \quad e^{2M_j + S_j^2} (e^{S_j^2} - 1).$$

Furthermore, the long-run variance of the size distribution of establishments is decreasing in α_j .

The assumption of lognormal shocks is arguably quite strong. Also, in practice, the empirical variance of establishments in a sector is quite sensitive to the measured size of the largest establishments and, hence, to measurement error in their sizes. This should not be surprising given that, by Proposition 3, for sectors with small physical capital shares the size distribution of establishments can be close to a Pareto distribution with shape coefficient one, for which the variance diverges. This suggests we should look at other measures of dispersion in the size distribution.

The following series of propositions characterizes the invariant distribution of establishment sizes for the class of probability distributions for A with compact support, and presents a different measure of dispersion in the size distribution. In particular, the 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 establishment sizes are measured relative to trend (or equivalently that population growth is zero) is sufficient to guarantee that establishment sizes lie in this compact set:

$$(13) \quad \ln n_{ij} \in LN \equiv \frac{\beta_j(1 - \alpha_j)}{(1 - \omega_j)(1 - \beta_j(1 - \alpha_j))} \times [-\ln \bar{A}, -\ln \underline{A}].$$

Under this assumption, we get the following proposition.

PROPOSITION 5 (Existence and Uniqueness): *If log productivity levels are bounded, then for any $\alpha_j, \beta_j, \omega_j \in (0, 1)$ there exists a unique invariant distribution of establishment sizes in sector j .*

We also want to establish how the size distribution of establishments in a sector varies with the physical capital share of a sector. Clearly, for any $\alpha_j, \beta_j, \omega_j \in (0, 1)$, the invariant distribution of establishment sizes has compact support and hence necessarily has thinner tails than the Pareto distribution which has an unbounded support. Next, we define an ordering over distributions with a common compact support in terms of the thinness of their tails, and show that sectors with larger physical capital shares have thinner tails according to this ordering. We make these notions precise in the following definition and proposition.

DEFINITION 6: *Let λ and ψ be probability measures on $[\underline{b}, \bar{b}]$. The probability measure λ has **thinner tails** than ψ if there exist \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 $\bar{x} \leq x \leq \bar{b}$, $\lambda([\bar{x}, x]) \leq \psi([\bar{x}, x])$.*

All other things equal, it is easy to see from equation (13) that an increase in the capital share produces a distribution with thinner tails as the support of the distribution shrinks. We can also prove a stronger result if we standardize the support of the size distribution produced by our model. This is also necessary to contrast the implications of our model with the data in which the size categories are the same for all

industries. Specifically, if we scale the productivity process A_{ij} by

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

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

PROPOSITION 7 (Thinner Tails): *For any $\alpha_j, \beta_j, \omega_j \in (0, 1)$, the invariant distribution of establishment sizes has thinner tails than the Pareto distribution with coefficient one. Other things equal, if $\alpha_j > \alpha_k$, then the invariant distribution of establishments in sector j has thinner tails than the one in sector k .*

So far we have established that the process of accumulating industry-specific human capital alone is sufficient to generate many properties of establishment size dynamics and distributions. In Appendix A, we demonstrate that these implications are also robust to relaxing a number of the modelling assumptions used above. We view the robustness of these results as a strength of our approach, which is particularly important given the robustness of the facts about establishment size dynamics. Another strength of our approach is that, since our theory uses the accumulation of industry-specific human capital to explain scale dependence, it also has new testable implications for the way in which the degree of scale dependence varies with the physical capital intensity of an industry. In the next section we examine these predictions using US data.

II. Evidence on Scale Dependence

Our model can reproduce the fact that establishment growth and net exit rates decline with establishment size, and that the size distribution of establishments has thinner tails than the Pareto distribution with shape coefficient one. More importantly, our model also has strong predictions for the way in which scale dependence varies across sectors. In our theory, the degree of reversion to the mean in human capital stocks, and therefore in establishment sizes, increases with the degree of diminishing returns

in human capital. As our model endogenously produces constant returns to scale at the industry level, large capital shares imply small human capital shares, and are evidence of a greater degree of diminishing returns in human capital. Hence, our model implies that if we look at sectors with large capital shares, we should observe a greater degree of scale dependence in growth rates and in net exit rates, and a size distribution of establishments with thinner tails. In this section, we show that this prediction is supported by results from a new dataset on establishment dynamics and size distributions for the private nonfarm US economy. This dataset is novel in that it provides data for a wider range of sectors and industries, and at a finer level of disaggregation, than previously available. We thus begin by first verifying these facts for the entire US economy. We then turn to an examination of the sectoral predictions of our model.

A. Data Sources and Economy-Wide Scale Dependence

We investigate scale dependence and its variation across sectors using data on growth rates, exit and entry rates (and so net exit rates), and the distribution of establishment sizes. We use two datasets constructed especially for this project by the US Census Bureau. For our purposes, these new datasets have several advantages over the publicly available data sources. One advantage is that they provide the number of establishments per size category for the smallest size categories that the US Census will release (given its confidentiality restrictions). Due to our emphasis on the tails of the size distribution, this level of detail is crucial, as previous analyses of the size distribution of establishments have used data for much coarser size categories. Second, these datasets include all sectors in the private nonfarm US economy, including both manufacturing and services. This is important for our study because we want to understand the effect of sectoral differences in physical capital shares on the size distribution of establishments, and the variation in physical capital shares is much larger across services and manufacturing sectors than within them. Third, the datasets refer to establishment sizes, and not enterprise sizes, which, as we have argued, is a better fit for our theory.

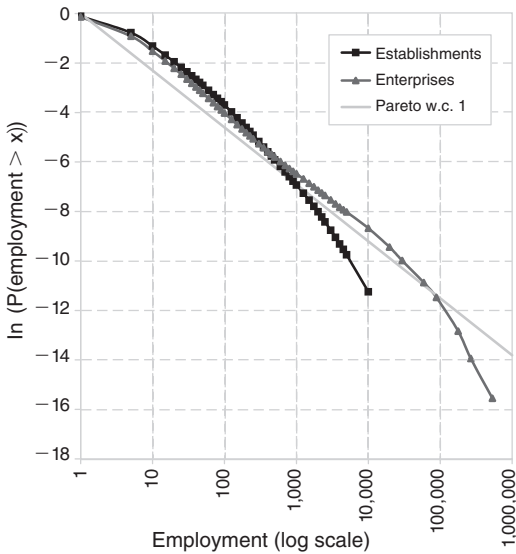


FIGURE 1. DISTRIBUTION OF ESTABLISHMENT AND ENTERPRISE SIZES IN 2000

Notes: The figure presents the probability that establishments and enterprises are larger than a particular size against that size in 2000. The figure also presents the same probability for a Pareto density with coefficient one. The data on enterprises are aggregated into 50 bins and into 43 bins for establishments.

Source: US Census Bureau, Statistics of US Businesses.

Cross-Section Data and the Size Distribution.—The first dataset is drawn from the Statistics of US Businesses (SUSB) program and contains data on establishment size distributions by sector at the two-digit SIC level for 1990 and three-digit NAICS level for 2000. The data are constructed from several sources, including the annual County Business Patterns (CBP) data files. Figure 1 illustrates the scale dependence in the size distribution of establishments by comparing the distribution 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 (for example, Robert L. Axtell 2001). For the Pareto distribution, the logarithm of the share of production units greater than a particular employment size varies linearly with the logarithm of employment, at a negative rate determined by its shape coeffi-

cient. Consequently, we plot the logarithm of the share of production units greater than a particular employment size against size for the actual size distributions.

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. In the figure, however, both look concave, reflecting the fact that they have thinner tails than the Pareto benchmark. As the Pareto distribution is scale independent, in the sense that the distribution is invariant to truncation of the left tail, this is evidence of scale dependence in the size distributions of both establishments and enterprises. It is worth noting that the size distribution of enterprises is much closer to the Pareto, especially if we focus attention on enterprises with between 50 and 10,000 employees. The differences between the size distributions for establishments and enterprises may shed light on the forces that determine the boundaries of the firm. Our theory focuses, however, on the technology of a single production unit and does not address questions of ownership or control. Consequently, this topic, although fascinating, is beyond the scope of this study, and hereafter we focus solely on establishment data.

Longitudinal Data, Growth, and Net Exit.—

To examine establishment size dynamics, we use a second new dataset drawn from the US Census Bureau's Business Information Tracking Series (BITS), which includes data on growth rates of establishments between 1990 and 2000 and deaths and births of establishments by size category for 1995–1996. The unique aspect of this longitudinal dataset is that it tracks the size of establishments for several years and, for exiting/entering establishments, for three years before/after they exit/enter.

With these data, we examine the well-known stylized fact that small establishments grow faster than large establishments, when attention is restricted to those establishments that remain in operation. This is illustrated in Figure 2, which plots growth rates by establishment employment for the United States over both one- and ten-year intervals. This figure shows that the difference in growth rates between small and large establishments can be as much as 20 percentage points within a year, and that the

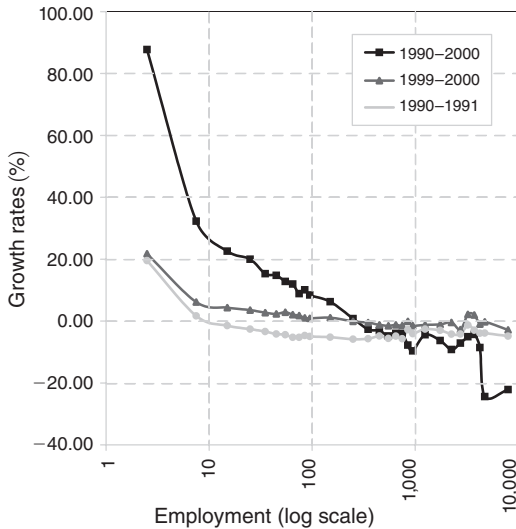


FIGURE 2. ESTABLISHMENT CONDITIONAL GROWTH RATES, 1990-2000

Notes: The figure presents average establishment employment growth rates by size bin for establishments that existed between 1990 and 2000, 1999 and 2000, and 1990 and 1991. Employment sizes are divided into 29 size bins. *Source:* US Census Bureau, Business Information Tracking Series.

accumulated effect of this pattern over a decade leads to differences of more than 100 percentage points between small and large establishments. Moreover, this scale dependence in growth rates is not limited to the smallest establishments and is significant throughout the size distribution. Note that Figure 2 presents data on establishments that survived the relevant period; hence, selection may be a relevant force explaining the exhibited scale dependence. (We will address this in detail below.)

Figures 1 and 2 show a significant degree of scale dependence in the size distribution and growth rates of establishments in the US economy. Below, we show that this scale dependence also manifests itself in net exit rates. However, this scale dependence has been documented using data aggregated by size category. Our theory predicts that we should also observe it for average industry (or representative) establishment sizes within a sector. As we have argued, our notion of an industry is very narrow, since it includes only establishments that produce the

same goods and use exactly the same technology and physical and human capital. We do not have data disaggregated at this level. Hence, hereafter we interpret each establishment in our dataset as a representative establishment in a narrowly defined industry which has the number of establishments per industry given by our theory.⁷

Selection, Survival, and Age Effects.—The theory outlined in Section II makes specific predictions for the growth rate of establishments, *conditional on their survival*. It also makes predictions about the behavior of the net exit rate of establishments and about the size distribution of establishments. Consequently, in the empirical analysis below, we focus, separately, on conditional growth rates, net exit rates, and size distributions. The focus on conditional growth rates contrasts with the empirical literature testing Gibrat's law, which has emphasized establishment growth rates not conditioned on survival and, in particular, the role of exit in reducing the unconditional growth rate of small establishments. We do not take that approach here for a number of reasons. One is that our theory makes specific predictions both for growth rates conditional on survival and for net exit rates, so we examine both directly. Also, the implications of all theories are sensitive to the precise way in which the growth rate of exiting establishments is treated, and whether entering establishments are also included. Moreover, there is no clear consensus as to the appropriate way to include entry; note the alternative empirical methodologies of Dunne, Roberts, and Samuelson (1989a, b) and Steven J. Davis and John Haltiwanger (1999). The theory of our work here continues to predict scale dependence under either of these methodologies. However, the fact that the same mechanism causes the scale dependence in conditional growth rates and net exit rates means that there exist yet further treatments of entry and exit that result in unconditional growth rates that display no scale dependence. This leads to the third reason for our conditional approach: by focusing on these facts separately, we can directly examine whether the degree of scale

⁷ A theoretically consistent empirical decomposition between industry and establishment heterogeneity requires unit record data which are not available for a broad sample like ours.

dependence in both conditional growth rates and net exit rates varies across sectors, as our theory predicts.

This focus also further distinguishes our approach from studies that emphasize selection mechanisms in producing scale-dependent growth. Although we acknowledge that selection effects may be important for small establishments, we interpret the evidence as suggesting that they are less so for the scale dependence observed across medium-sized and large establishments. For example, one important prediction of selection theories is that establishments should become smaller in the years prior to exit, which is often referred to as the “Shadow of Death” (Zvi Griliches and Haim Regev 1995). However, our gross exit data show little evidence of the Shadow of Death; that is, most establishments do not appear to decline in size in the years leading up to their exit. The only declines are for very small establishments and particularly for those that survive fewer than three years. This suggests that selection may be important for small, young establishments, but not for medium-sized and large ones. In contrast, our theory predicts that establishments will exit and enter at all sizes, as the equilibrium number of establishments in an industry adjusts to productivity shocks.

Our theory emphasizes the role of size in establishment dynamics but, as many theories do, it abstracts completely from age effects. In our theory, young establishments behave identically to old establishments: size, but not age, matters. Unfortunately, our dataset does not contain information on age and so we are not able to present results for given age cohorts. Therefore, some of the scale dependence we document may be the result of age effects. However, the preceding empirical literature on establishment dynamics has found that scale effects are important even after controlling for age (for example, Evans 1987a, b; Hall 1987; and Davis and Haltiwanger 1999⁸). Moreover, the magnitude of the age dependence documented by Davis and Haltiwanger (1999) for the United States is much smaller than the scale dependence we

have found in our data.⁹ Age effects also seem to diminish quite quickly with establishment age, and the scale dependence in growth rates that we document and will use later concerns ten-year growth rates conditional on survival.

Sectoral Capital Shares.—To examine the implications of our model for cross-sectoral differences in mean reversion, we will need data on physical capital shares, which come from the US Bureau of Economic Analysis (BEA) industry accounts. We use the BEA data on labor costs and value added at basic prices to construct labor shares that include human capital. We then construct physical capital shares as one minus the labor share. This method implies that the physical capital shares we use include everything that is not classified as labor. This suggests two potential problems with our computations. One is that the physical capital shares include land shares. Land is not an industry-specific factor, but because its share is usually small, this should have a negligible effect on the physical capital shares we use. The other potential problem is that we are using the physical capital share in value added, but our theory abstracts from the use of intermediate inputs. To address the first problem, we focus on sectors with physical capital shares smaller than one-half, although the results are similar if we consider all sectors. To address the second problem, we also present results with physical capital shares adjusted for the share of value added and the share of materials purchased from the same sector.

B. Evidence on Sectoral Scale Dependence

Besides the economy-wide scale dependence just documented, our theory implies that scale dependence should be larger in sectors that use physical capital more intensively. This implication distinguishes our theory from other theories that may also imply economy-wide scale dependence. For example, theories that emphasize financial constraints in explaining

⁸ Davis and Haltiwanger (1999) study only unconditional growth rates.

⁹ Using more aggregated data for enterprises with only one establishment, we can verify that the predictions of our theory for sectoral variation in the thickness of the tails of the size distribution hold for establishments younger and older than five years in the leading example used below.

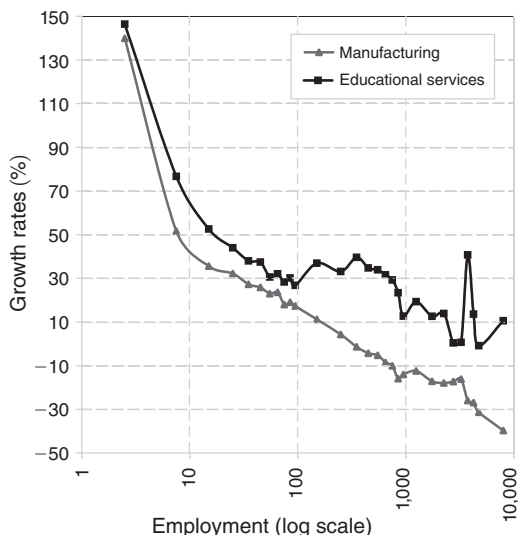


FIGURE 3. ESTABLISHMENT CONDITIONAL GROWTH RATES BY SECTOR, 1990–2000

Notes: The figure presents average establishment employment growth rates by size bin for establishments that existed between 1990 and 2000 in the educational services and manufacturing sectors. In both sectors, employment sizes are divided into 29 size bins.

Source: US Census Bureau, Business Information Tracking Series.

scale dependence predict that scale dependence should be more pronounced in sectors in which establishments have less collateral. This plausibly corresponds to sectors in which the human capital share is relatively large and the physical capital share relatively small, which is the opposite prediction of our theory. We now present evidence on the sectoral variation in scale dependence for conditional growth rates, net exit rates, and size distributions, and we show that this evidence corroborates the implications of our theory.

Growth Rates of Surviving Establishments.— We begin by examining 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 share of 0.397. If our theory is consistent with the data, we should see growth rates of manufacturing establishments decline faster with size than growth rates of establishments in the edu-

cational services sector (Proposition 1). That is what we see in Figure 3. In the figure, the differences are very large over a period of ten years (1990–2000). Not only do small establishments grow faster than large establishments in both sectors, but the scale dependence is significant for the entire range of establishment sizes. The difference between the growth rates in these two sectors increases with establishment size and is, for the largest establishments, more than 40 percentage points.

This evidence is not peculiar to the pair of sectors in the example. We demonstrate this by examining the same implication of our theory for all sectors. We use data on the growth rate of establishments, n_{t+1j}/n_{tj} , in a particular size category, n_{ij} , and estimate the regression specified by equation (11):

$$(14) \quad \ln\left(\frac{n_{t+1j}}{n_{tj}}\right) = \tilde{a}_j + \tilde{b} \ln n_{ij} + \tilde{c}\alpha_j \ln n_{ij} + \tilde{\epsilon}_{ij},$$

where

$$\tilde{a}_j = n_j^c,$$

$$\tilde{b} = -(1 - \omega_j)(1 - \beta_j),$$

$$\tilde{c} = -(1 - \omega_j)\beta_j, \text{ and}$$

$$\tilde{\epsilon}_{ij} = -\beta_j(1 - \alpha_j) \ln A_{t+1j}.$$

Notice that a full structural estimation of our model would require \tilde{b} and \tilde{c} to vary as β_j and ω_j vary by sector. Unfortunately, we do not have data on the share of industry-specific human capital in labor services or the share of investment in human capital production, and so we assume that these two shares do not vary across sectors or, if they do, that they are uncorrelated with capital shares. If they are uncorrelated, then our estimation strategy is not efficient, but the coefficients are still unbiased and consistent. Given that all the results presented below are significant at a 1 percent level, the lack of efficiency of the estimator is not worrisome.¹⁰

¹⁰ If β_j and ω_j are correlated with α_j , then the estimates of \tilde{c} are biased. Let $\beta_j = \beta + \beta_j^e$, where β is the mean of β_j . Similarly, let $\omega_j = \omega + \omega_j^e$. Then, the sign of the bias depends on the sign of the covariance between $\omega_j^e \beta + \beta_j^e(1 - \omega) + 2\omega_j^e \beta_j^e$ and α_j . If $\omega_j^e = 0$ for all j , then if $\text{Cov}(\beta_j^e, \alpha_j) \geq 0$,

TABLE 1—ESTABLISHMENT GROWTH AND CAPITAL INTENSITY

	Growth rates 1990–2000			
	Variance = $1/\mu_j$		Variance = $(1-\alpha_j)^2/\mu_j$	
	(Adjusted)		(Adjusted)	
$\tilde{\epsilon}$	-0.1115	-0.1517	-0.1488	-0.1814
Standard error	0.0255	0.0314	0.0304	0.0325
<i>p</i> -value	0.0000	0.0000	0.0000	0.0000
# of Observations	940	1,082	940	1,082

Notes: This table presents the estimates of $\tilde{\epsilon}$ obtained from the regression equation specified in (14). Estimates are presented under two different assumptions on the form of heteroskedasticity as described in the text.

Source: US Census Bureau, Business Information Tracking Series.

Our estimation procedure lets us back out the average level of these parameters for the United States (which we compute below). Apart from this caveat, our empirical exercise uses precisely the structure imposed by our model.

A requirement for a structural estimation of our model is to account for the heteroskedasticity that the model implies. In particular, if the variance of technology shocks is constant across sectors, the model predicts that the variance of the error term decreases with α_j . We use generalized least squares (GLS) to take this effect into account and estimate equation (14) with and without the effect of physical capital shares on the variance.

We estimate this relationship using GLS to take into account the fact that the smaller-size categories have many more establishments, as well as the heteroskedasticity predicted by our model. Of course, taking account of heteroskedasticity should improve the efficiency of our estimation, but the results without doing so are still unbiased and consistent. We calculate the

our estimates are biased toward zero. If true, the estimates below are a lower bound for the true effect, which strengthens our conclusions. If $\beta_j^e = 0$ for all j , then if $\text{Cov}(\omega_j^e, \alpha_j) \geq 0$, we again get estimates biased toward zero, which again reinforces our results. In general, in order for our estimates to be biased toward zero, we need to assume that $\text{Cov}(\omega_j^e \beta_j + \beta_j^e(1 - \omega) + 2\omega_j^e \beta_j^e, \alpha_j) \geq 0$. This restriction amounts to saying that human capital depreciates more slowly, and accounts for a larger share, in industries that are more capital intensive. This would be the case if, for example, industry-specific human capital is needed to operate industry-specific machines and if those machines depreciate more slowly in industries that use more physical capital. We leave this as a restriction imposed on the empirical model.

weights using data on the number of establishments in each size category and physical capital shares. Throughout, we use average US physical capital shares for the period 1990–2000. The theory predicts that the estimate of $\tilde{\epsilon}$ should be negative and significant.

The estimates of $\tilde{\epsilon}$ are presented in Table 1. The first and third columns present the estimates using average capital shares for the period 1990–2000. The first two columns weight observations only by the number of establishments in a particular size class bin. The last two columns adjust also for the heteroskedasticity predicted by our model. Although the estimates of $\tilde{\epsilon}$ are negative and strongly significant in all regressions, the results are the strongest when we use the exact specification given by our theory, namely, variance terms that depend on capital shares. Given the largest establishment size in our sample, a larger (in absolute value) coefficient implies more scale dependence for all establishment sizes. The results in Table 1 show that scale dependence increases significantly with sectoral physical capital shares: a doubling in the size of establishments in manufacturing ($\alpha_j \approx 1/3$) decreases average growth by about 5 percent relative to the educational services sector ($\alpha_j \approx 0$).

As mentioned above, the physical capital shares have been calculated as one minus the share of labor costs in value added. Since materials are a large fraction of gross output in an industry, however, 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 it empirically, we

also 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 multiplied this share by the original physical capital share to obtain an *adjusted* physical capital share. If all intermediate inputs originated in the same sector, then the original physical capital shares would equal the adjusted physical capital shares. If the rest of the materials used in production are homogeneous, then 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. The omission of these effects could, however, be expected to bias our coefficients toward zero. Given the statistical significance of our results presented in the second and fourth columns of Table 1, we do not think the omission undermines our empirical strategy.¹¹ The omission of intermediate inputs from other sectors may account for some of the unexplained variation in growth rates; so may 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.

Our estimation of \tilde{b} and \tilde{e} assumes that both β_j and ω_j are constant, or independent of α_j , across sectors. (Call the average or constant values β and ω , respectively.) We can then use the estimates presented in Table 1, together with the estimates of \tilde{b} and equation (11), to infer values for β and ω . The estimates of \tilde{b} for the exercises in the first two columns of Table 1 are -0.146 (s.e. 0.009) and -0.154 (s.e. 0.008). When we use the capital share in the weights in the third and fourth columns, they become -0.134 (s.e. 0.011) and -0.145 (s.e. 0.008). These values imply a share of industry-specific human capital in labor services between 0.432 and 0.556 (where $\beta = 1/(1 + \tilde{b}/\tilde{e})$). That is, the model and the estimation above imply that the share of labor services related to industry-specific human capital is roughly half. Thus, as we have argued, the share of industry-specific human capital consistent with the scale depen-

dence in establishment dynamics is very significant. Other forms of human capital that are not industry-specific (those associated with individuals rather than industries) are, of course, not included in this share. Our estimates also imply an average share of investment in human capital production ($1 - \omega = \tilde{e}/\beta$) between 0.258 and 0.326. This investment share depends on the period over which we calculate growth rates, which in this case is ten years.

The last ten years have witnessed a substantial decline in employment among large manufacturing establishments. A potential concern is that this phenomenon may be driving the larger-scale dependence observed in physical capital-intensive manufacturing sectors. To address this concern, we replicate the previous exercise for manufacturing and nonmanufacturing sectors separately. The results presented in Table 2 show that this phenomenon is not driving the results in Table 1. The point estimates for nonmanufacturing are close to those for the whole economy and are strongly significant. For manufacturing, the estimates are much less precise, reflecting the smaller variation in physical capital shares among these sectors. This was precisely our justification for using all sectors in the economy.

Net Exit Rates.—Our mechanism, which emphasizes mean reversion in stocks of specific factors, when combined with an assumption on the level of the elasticity of substitution, also implies that net exit rates should decline with establishment size. Furthermore, the rate of decline should increase with physical capital shares.¹² Using BITS data for US sectors in 1995–1996, we run the following regression:

$$(15) \quad \ln(1 + NER_j) = \check{\alpha}_j + \check{b} \ln n_j + \check{\epsilon}\alpha_j \ln n_j + \check{\epsilon}_{ij},$$

where NER_j denotes the net exit rate in sector j

¹¹ Adjusting the physical capital shares increases the number of sectors in our sample with physical capital shares below one-half from 44 to 52.

¹² Orr (1974), Gorecki (1976), Hause and Du Rietz (1984), and MacDonald (1986) find that firm, not establishment, exit rates are 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.

TABLE 2—ESTABLISHMENT GROWTH AND CAPITAL INTENSITY
IN MANUFACTURING AND NONMANUFACTURING SECTORS

	Growth rates, 1990–2000			
	Variance = $1/\mu_j$			
	Manufacturing		Nonmanufacturing	
	(Adjusted)		(Adjusted)	
\tilde{e}	-0.0524	-0.0485	-0.1159	-0.1619
Standard error	0.0981	0.1213	0.0265	0.0329
<i>p</i> -value	0.5930	0.6900	0.0000	0.0000
	Growth rates 1990–2000			
	Variance = $(1 - \alpha_j)^2/\mu_j$			
	Manufacturing		Nonmanufacturing	
	(Adjusted)		(Adjusted)	
\tilde{e}	-0.0876	-0.0720	-0.1556	-0.1922
Standard error	0.0972	0.1295	0.0322	0.0342
<i>p</i> -value	0.3680	0.5780	0.0000	0.0000
# of Observations	940	1,082	940	1,082

Note: This table presents the estimates of \tilde{e} obtained from the regression equation specified in (14) estimated on the manufacturing and nonmanufacturing sector subsamples.

Source: US Census Bureau, Business Information Tracking Series.

and size bin n_j . This amounts to estimating the exponential relationship between net exit rates and sizes implied by the model. The results are presented in Table 3 for net exit rates in 1995–1996.¹³ The top panel shows the results using GLS, where the weights in the variance-covariance matrix include only the number of establishments.¹⁴ The bottom panel presents results of the same exercise with the weights adjusted to take into account the heteroskedasticity predicted by the theory. Again, we present all results using average capital shares for the period 1990–2000, along with results using adjusted capital shares. We also present the results if we measure establishment size one year before/after exit/entry.

The results are consistent with our theory: all the estimates are negative and significant.

¹³ The results are very similar if we use net exit rates in 1993–1994.

¹⁴ The measure of size we use is given by $\tilde{\mu}_j = ((\mu_{jy_1} + \mu_{jy_2})/2)^2 / (\mu_{jy_1} + E_{jy_1})$, where μ_{jy_2} is the number of establishments in year y_2 of a given size indexed by j , and E_{jy_1} is the number of establishments that entered in y_2 of a given size j . The reason we use this measure is that, in contrast with the growth rate regressions, we should not just use the number of surviving firms, but rather the sum of all firms alive before exit and after entry.

The results are also economically significant. As shown in Table 3, a doubling of establishment size decreases net exit rates by around 1 percent ($\approx -0.0314 \times 0.33$) in manufacturing, while net exit rates decline little with size in educational services. The results are still very significant, but smaller in magnitude when we measure size one year before/after entry/exit.

Finally, note that our theory, with logarithmic preferences, implies that the estimates of \tilde{e} should be identical to the estimates of \tilde{e} for conditional growth rates, as presented earlier (Table 1), once we correct for the fact that the share of investment in human capital accumulation $(1 - \omega)$ should be smaller for one year (as for net exit rates) than for ten years (as for growth rates). We can use one-year growth rates in 1990 and net exit rates in 1993–1994 (using size also in 1993–1994) and estimate the two equations jointly, restricting the parameters to be the same. (The difference in years is due to data limitations.) Now, we get estimates for $\tilde{e} = \tilde{e}$ of -0.0507 (s.e. 0.006), with unadjusted capital shares, and -0.0418 (s.e. 0.007) with adjusted capital shares. If we take into account the heteroskedasticity implied by the model, we get -0.0496 (s.e. 0.007) and -0.0372 (s.e. 0.008), respectively. Hence, restricting the parameters

TABLE 3—NET EXIT RATES AND CAPITAL INDUSTRY

Net exit rate, 1995–1996 $\left(\frac{\text{Exit}_{96} - \text{Entry}_{95}}{(\# \text{ Establishments}_{96} + \# \text{ Establishments}_{95})/2} \right)$	Variance = $1/\hat{\mu}_j$			
	Size in 1995–1996		Size in 1994–1997	
	(Adjusted)		(Adjusted)	
$\check{\epsilon}$	–0.0314	–0.0331	–0.0172	–0.0186
Standard error	0.0029	0.0034	0.0024	0.0028
<i>p</i> -value	0.0000	0.0000	0.0000	0.0000
	Weights = $(1 - \alpha_j)^2/\hat{\mu}_j$			
	Size in 1995–1996		Size in 1994–1997	
	(Adjusted)		(Adjusted)	
$\check{\epsilon}$	–0.0324	–0.0280	–0.0164	–0.0151
Standard error	0.0036	0.0036	0.0029	0.0030
<i>p</i> -value	0.0000	0.0000	0.0000	0.0000
# of Observations	1,733	2,029	1,721	2,012

Notes: This table presents the estimates of $\check{\epsilon}$ obtained from the regression equation specified in (15). Estimates are presented under two different assumptions on the form of heteroskedasticity as described in the text.

Source: US Census Bureau, Business Information Tracking Series.

to be the same yields results that are very much in line with the results in Table 3. This confirms yet another of the theory's predictions: that the amount of scale dependence in both growth and net exit rates should be similar within a given sector.

Size Distributions.—We now turn to the implications of our theory for the size distribution of establishments. As we have argued above, within a sector, the variance of growth rates and net exit rates, and therefore establishment sizes, should be smaller the larger the sector's physical capital share, given a common volatility of shocks across sectors.¹⁵ In fact, in some exercises above, we have used this prediction to calculate efficient estimates of the variation of scale dependence with physical capital shares. It is natural, therefore, to look directly at the relationship between the standard deviation of establishment sizes and physical capital shares across sectors. In Figure 4, we plot the standard deviation of establishment sizes against adjusted

physical capital shares for the years 1990 and 2000, together with fitted lines for both years. As the theory predicts, the relationship between the standard deviation and the adjusted physical capital shares is negative. It is also similar for both years.¹⁶

In Section ID, we also examined an alternate measure of the amount of dispersion in establishment sizes based on the thinness of the tails of the size distribution of establishments. This measure has the advantage of being less sensitive to the size of the largest establishments in the sample, which is particularly important for sectors in which the size distribution is well approximated by a Pareto distribution for which second moments are not defined. To compare thinness of tails, we use data from SUSB to calculate the share of establishments in sector j with employment larger than n_j , denoted by P_j . If the distribution of establishment sizes is Pareto

¹⁵ We do not have evidence that suggests that the volatility of productivity shocks is common across sectors. Differences in the underlying distribution of productivity shocks may result in difference in the standard deviation of establishment sizes not explained by our mechanism.

¹⁶ The fitted lines are calculated by running an OLS regression of the standard deviation of establishment sizes by sector against capital shares. The estimated coefficient is significant at the 1 percent level. Even if we eliminate some of the industries that seem to be outliers in Figure 4 (excluding from the regression the five industries with lowest capital shares and the five industries with highest capital shares), the estimated coefficient remains significant at the 1 percent level.

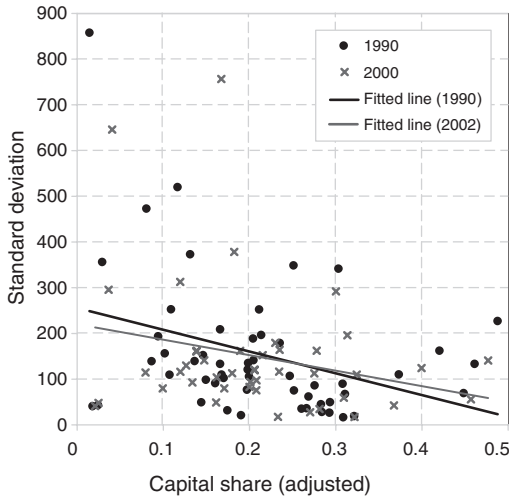


FIGURE 4. STANDARD DEVIATION OF ESTABLISHMENT SIZES AND CAPITAL SHARES, 1990 AND 2000

Notes: The figure presents the variance of establishment sizes by sector and the corresponding physical capital shares for 1990 and 2000. The variance is computed from data aggregated into size bins.

Source: US Census Bureau, Statistics of US Businesses.

with coefficient one, or growth rates are scale independent, then the relationship between $\ln P_j$ and $\ln n_j$ should be linear with slope minus one. If growth rates depend negatively on scale, then 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 to physical capital shares (Proposition 7).

A first look at the data confirms that prediction. In Figure 5, we plot $\ln P_j$ and n_j for the manufacturing and educational services sectors. This representation of the size distribution emphasizes the degree of concavity and makes differences between the two distributions particularly clear. The distribution of establishment sizes in the educational sector has more mass for very small and large establishments and less mass for intermediate establishments than the distribution in the manufacturing sector. This is particularly clear for small establishments. The figure also compares these distributions with the Pareto distribution with coefficient one (which corresponds to a straight line with slope -1 in Figure 5). The Pareto distribution with coefficient one has even more mass at the tails and

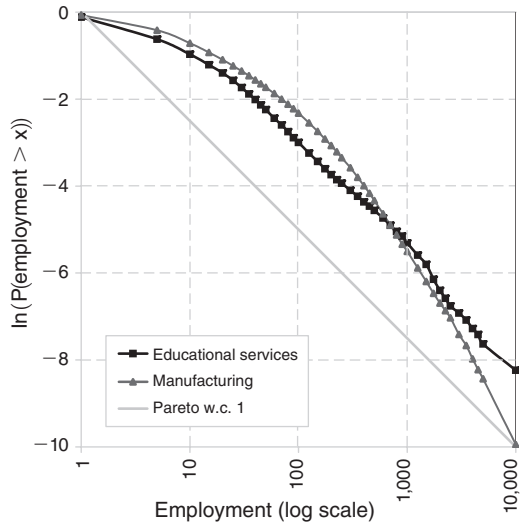


FIGURE 5. DISTRIBUTION OF ESTABLISHMENT SIZES BY SECTOR, 2000

Notes: The figure presents the probability that establishments in the educational services and manufacturing sectors are larger than a particular size against that size in 2000. It also presents the same probability for a Pareto density with coefficient one. The data on the number of establishments are aggregated into 43 bins.

Source: US Census Bureau, Statistics of US Businesses.

less at the center, consistent with Proposition 7 as long as $\beta_j, \omega_j, (1 - \alpha_j) > 0$. Both sectors have thinner tails than the Pareto, but as the theory predicts, the difference is larger for the manufacturing sector. Moreover, the differences between these distributions are economically large: in order to transform the size distribution of the manufacturing sector to that of the educational services sector, about 20 percent of the labor force that currently works in medium-sized manufacturing establishments would need to be reallocated to establishments with fewer than 50 or more than 1,000 employees.

In order to test the relationship between physical capital shares and the size distribution of establishments for all sectors, we use our new dataset on the size distributions of establishments for 1990 and 2000. We estimate the following regression:

$$(16) \quad \ln P_j = \hat{a}_j + \hat{b}_j \ln n_j + \hat{d} (\ln n_j)^2 + \hat{e}\alpha_j (\ln n_j)^2 + \hat{\epsilon}_j,$$

TABLE 4—SIZE DISTRIBUTIONS AND CAPITAL INTENSITY

Size distribution	1990		2000	
	(Adjusted)		(Adjusted)	
$\hat{\epsilon}$	-0.1015	-0.0402	-0.0730	-0.1309
Standard error	0.0152	0.0145	0.0167	0.0163
<i>p</i> -value	0.0000	0.0060	0.0000	0.0000
# of Observations	1,864	2,182	1,486	1,799

Note: This table presents the estimates of $\hat{\epsilon}$ obtained from the regression equation specified in (16).

Source: US Census Bureau, Statistics of US Businesses.

where \hat{a}_j and \hat{b}_j are industry-specific coefficients. This amounts to fitting a second-order equation and constraining the quadratic term to vary linearly with the physical capital share. The model now predicts that $\hat{\epsilon}$ should be negative and significant.¹⁷ The results are presented in Table 4, which shows that the estimate of $\hat{\epsilon}$ for both 1990 and 2000 is negative and strongly significant. The physical capital shares used in the regression are those corresponding to 1990 and 2000. The results with the adjusted physical capital shares further confirm the empirical significance of the mechanism in our theory.

III. Conclusion

In this paper, we have constructed a theory that is consistent with some well-known facts on scale dependence in establishment size dynamics and distributions. The theory emphasizes the role of the accumulation of industry-specific human capital and is robust to institutional and economic differences across sectors. We claim that the ubiquitous presence of these facts has to be the result of a mechanism, like ours, that is general enough to be 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 capital, 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 establishment growth rates, net exit rates, and distributions. We take this prediction to the data and show that it is a surprisingly good description for the cross section of US sectors. Any theory of establishment size dynamics has to confront this new evidence.

Other studies have emphasized various forms of market inefficiencies, such as financial frictions, as explanations of scale dependence. Financial frictions are also often cited as justification for various forms of policy intervention such as subsidies to small business. What we have shown is that, even though these frictions may be important for very small establishments, they are not necessary to explain the degree of scale dependence and its variation across sectors. As a consequence, if policy is to be applied to address these frictions, it is important that it be designed so as not to interfere with the growth and net exit of larger and existing establishments, which are well described by our efficient theory.

Our findings could also help assess differences in the efficiency of resource allocation across countries. Nezh Guner, Gustavo Ventura, and Xu Yi (2005) and Diego Restuccia and Richard Rogerson (2004) have argued that government policies targeted at establishments of specific sizes may explain large cross-country differences in productivity. International evidence on establishments dynamics and the size distribution of establishments, when combined with our efficient benchmark, could shed some light on the empirical significance of such scale-dependent policies. Other research (see, for example, the survey by James R. Tybout (2000), and the empirical evidence presented by Leo Sleuwaegen and Micheline Goedhuys (2002))

¹⁷ This prediction is the result of Proposition 7. Since Proposition 7 does not provide an explicit expression for the distribution of establishments, we estimate the variation in the thinness of the tails as a function of capital shares using the quadratic equation in the text.

has concluded that the thicker tails of the size distribution of establishments in sub-Saharan African countries, relative to that in the United States, are evidence of corruption in these countries; establishments either stay small in order to avoid official notice, or grow until they are large enough to co-opt the system to their own benefit. Our empirical finding that the size distributions in sectors with lower physical capital intensities display thicker tails may strengthen this conclusion, given the greater concentration of the US economy in human capital-intensive sectors.

Finally, by emphasizing the accumulation of specific human capital, our theory makes

predictions, as well, for the evolution of the establishment size distribution. The ongoing specialization of developed economies in services will lead to a more dispersed distribution of establishment sizes, where we will see more small and more very large establishments. This argument suggests that we are moving toward an economy in which the dominance of large establishments in some industries, like Walmart stores, will coexist increasingly with large numbers of small establishments in different industries within the same sector, like bakeries or tailors. According to our analysis, this trend is the natural result of the efficient division of an industry's production among establishments.

APPENDIX A: ROBUSTNESS OF THE MECHANISM

We have argued that any proposed explanation for the documented patterns in establishment size dynamics and distribution must be robust to the wide variety of differences in institutions and market structures for which these patterns have been observed. Here, we establish that the mechanism described above is preserved as we generalize the model in a number of ways.

Establishment Costs.—In the model above, establishments face decreasing returns to scale and a fixed cost denominated in terms of the establishment's output. This combination implies that the output of the establishment is constant, so that establishments reduce employment (and hence size in terms of employment) when the stock of specific human capital grows: reversion to the mean in the stock of specific factors *from above* produces reversion to the mean in establishment sizes *from below*. Changes in the specification of the cost structure can reverse the qualitative relationship between factor supplies and establishment size, but do not change our basic results.

To see this, assume as before that each establishment in industry j in period t produces output according to equation (6). Now, however, assume that fixed costs depend on the average number of workers, \bar{n}_{tj} , and 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 establishments in the industry. Individual establishments do not take into account the effect of their hiring decisions on the fixed costs, so the problem of the establishment is identical to the one presented above. We assume that $0 \leq \xi_j < 1$, so if $\xi_j = 0$ we have exactly the same case studied in the text. Taking first-order conditions and allowing for free entry and exit, we find that

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

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

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

This equation is analogous to one derived before, with the difference that both employment and output respond to changes in factor supplies.¹⁸ Moreover, the *direction* of the change can differ: if $\xi_j > \gamma_j$, increases in the stock of industry-specific physical and human capital lead to *increases* in the size of the establishment. As long as the size of an establishment responds monotonically to the stock of human capital, mean reversion in factor supplies will drive mean reversion in establishment sizes, and the main properties for establishment growth and exit rates, and the size distribution, are preserved. The direction of this response is not important: when $\xi_j > \gamma_j$, reversion to the mean in the stock of specific factors *from above* produces reversion to the mean in establishment sizes *from above*.

Within-Industry Establishment Heterogeneity.—In our theory above, we have abstracted from heterogeneity among establishments *within* an industry in order to focus attention on heterogeneity *across* industries. Yet, differences in establishment sizes exist even within narrowly defined industries. Now we show how such heterogeneity can be added without introducing selection effects or changing the key empirical implications of our theory.

Suppose that after having decided to produce in a period and paying the fixed cost F , each establishment $i \in [0, \mu]$ observes an establishment-specific productivity shock z_i , where we suppress time and industry subscripts. This shock is assumed to be i.i.d. over time, establishments, and industries within a sector. After observing this shock, establishment i hires labor n_i and industry- j -specific physical capital k_i and human capital h_i to produce output according to

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

It is straightforward to show that, in equilibrium, the relative levels of output and factors across establishments in an industry are given by

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

so that establishments with a larger shock use more of both inputs and produce more output. This leads to an industry-level production function that has exactly the same form as for the original problem. Consequently, the choices of N and μ , as well as investment in both types of capital, are analogous, and the implications for the *average* growth and net exit rates of establishments in that industry, or the *average* size distribution of establishments, are unchanged. Establishment-specific shocks will increase the amount of mean reversion within all sectors, but the relationship with factor intensities across sectors is maintained as long as the distribution of shocks within a sector is independent of factor intensities.

In more general models of within-industry heterogeneity, there will be an active selection mechanism, and this will lead to a survivorship bias as establishments with low productivity realizations exit. In such models, the estimated *level* of scale dependence will be affected by this bias. However, as long as this bias is not related to factor intensities, it will not affect *differences* in the degree of scale dependence across sectors. Under other plausible assumptions, for example if establishments in capital-intensive sectors are less likely to exit in response to a negative shock because of large setup costs, then the survivorship bias will be smaller for capital-intensive sectors and our empirical results are strengthened.

Learning-by-Doing Externalities.—So far we have assumed that human capital accumulation requires some industry-specific inputs. The dependence on industry-specific inputs is important for

¹⁸ Notice that because the fixed costs entail an external cost, the equilibrium will not be Pareto optimal. However, a pseudo-social planner problem can be set up that yields the same aggregate implications as the problem discussed in Section I.

our model; it allows human capital accumulation to vary with industry output and is the primary source of mean reversion at the industry level. The inputs to learning are purchased by consumers, and the resulting level of human capital is rented out by consumers, so that there is no externality. An alternative assumption that has similar effects is the assumption that human capital is accumulated from *learning-by-doing externalities*, so that

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

Thus, the larger is output in the industry, the larger is human capital accumulation. Importantly, this involves no resource cost to the economy. Suppose 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 that human capital operates exactly like labor-augmenting technological progress.

With this change, we can no longer use the social planners problem to solve for equilibrium allocations in this model, but we can use a pseudo-planner's problem. Similar reasoning then produces an expression for the normalized rate of growth of an establishment given by

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

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

This extension emphasizes that what matters for our results on mean reversion is not industry-specific human capital *per se*, but rather the sensitivity of current production decisions to past output in the industry. This is important in light of recent research by Kambourov and Manovskii (2005), who find little evidence in individual earnings data for the existence of industry-specific human capital.¹⁹ However, their result is consistent with industry-specific learning-by-doing externalities in which individual workers do not appropriate the returns to industry-specific human capital.

Monopolistic Competition.—The theory in the text uses a competitive model of establishments to derive conclusions on the growth, exit, and size distribution of establishments. Now we demonstrate that our conclusions are not sensitive to this specification for the organization of production in establishments by deriving similar conclusions from a Dixit-Stiglitz monopolistic competition model with a taste for variety in which we identify an establishment with each variety. In this model, substitution for varieties in the same industry limits demand for a particular variety in an industry and thus determines the size of the establishment. The industry expands and contracts through variation in the number of establishments.

We assume that each industry consists of a continuum of potential varieties, which we index by ϖ . 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. The problem of a consumer is to purchase goods and accumulate industry-specific capitals to maximize lifetime utility:

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

¹⁹ Our 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 (2005).

subject to

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

$$K_{t+1j} = K_{tj}^{\lambda_j} X_t^{1-\lambda_j}, \quad H_{t+1j} = A_{t+1j} H_{tj}^{\omega_j} I_{tj}^{1-\omega_j},$$

$$Q_{tj} + I_{tj} \equiv E_{tj} \leq \left\{ \int_{0 \leq \varpi \leq \Omega_j} (D_{tj\varpi})^{(\sigma_j-1)/\sigma_j} d\varpi \right\}^{\sigma_j/(\sigma_j-1)},$$

$$C_t + X_t = \prod_{j=1}^J (Q_{tj})^{\theta_j}, \quad \sum_{j=1}^J N_{tj} \leq N_t,$$

for all t and all j , where E_{tj} is total demand for the final good from industry j , $p_{tj\varpi}$ is the price of variety ϖ , 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.

An establishment produces a variety ϖ using a constant returns-to-scale Cobb-Douglas technology $y_{\varpi} = k_{\varpi}^{\alpha} [h_{\varpi}^{\beta} n_{\varpi}^{1-\beta}]^{1-\alpha}$. Profit maximization combined with free entry yields industry output given by

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

This function has constant returns to scale, with TFP given by a function of the elasticity of substitution. The size of establishments in terms of employees is

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

which is a similar expression to the one derived for a perfect competition version of the theory. As a result, this model has identical implications for establishment dynamics and the size distribution.

All that remains to calculate the dynamics of establishment sizes is to solve for the accumulation decisions of agents. The log-linearity of the model implies that the equilibrium allocations can be obtained as the solution of a pseudo-planner’s problem identical to the planner’s problem in the text, 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 Nancy L. Stokey, Lucas, and Edward C. Prescott 1989). The result is a process for the evolution of establishment sizes that parallels the evolution of establishment sizes in the model with perfect competition.

APPENDIX B: PROOFS

PROOF OF PROPOSITION 3:

See Rossi-Hansberg and Wright (2007) Proposition 3.

PROOF OF PROPOSITION 4:

As shown in the text, if we detrend the growth rate of surviving establishments, then the invariant distribution of representative or average establishment sizes, in logs, is normal with mean M_j and variance as in (12). If we let $y = \ln n$, then the distribution of actual establishment sizes has to be weighted by N_j/e^y . If we normalize the distribution to be a probability distribution, we obtain the pdf of the size distribution

$$\frac{1}{S\sqrt{2\pi}} e^{-[y-(M_j-S^2)]^2/2S^2},$$

which means that the actual size distribution of establishments is also lognormal with mean $\exp(M_j - S^2/2)$ and variance

$$e^{2M_j+S_j^2} (e^{S_j^2}-1).$$

That the variance is decreasing in α_j follows from the fact that both the mean and variance of representative establishment sizes are decreasing in α_j .

PROOF OF PROPOSITION 5:

The proof is independent for each sector, so we drop j from the notation. The size of an establishment in period $t + 1$ is given by

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

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 in the text. 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^* , which 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 a unique invariant distribution exists, if we can show that the transition probability function Q satisfies the Feller property, is monotone, and satisfies the mixing condition.

To see that Q 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, then $f(g(\cdot))$ is continuous and bounded, and therefore so is Tf . Hence, T maps the space of bounded continuous functions into itself, $T : C(\bar{S}) \rightarrow C(\bar{S})$. To see that Q is monotone, we need to prove that if $f : LN \rightarrow \mathbb{R}$ is a nondecreasing function, then so is Tf . But this follows from the fact that g is nondecreasing in n . Hence, $f(g(n, A))$

is nondecreasing in n , and, therefore, so is Tf . Finally, to show that Q satisfies the mixing condition, we need to show that there exist $c \in LN$ and $\eta > 0$ such that

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

and

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

Now let $c = 0$. Since g is continuous and decreasing in A , there exists an A' such that for all $A \leq A'$, $g(n, A) > 0$. Then let $\eta' = 1 - \mu(A')$. Similarly, there exists an A'' such that for all $A \leq A''$, $g(n, A) < 0$. Finally, 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) guarantees that a unique invariant distribution exists, and that the iterates of T^* converge weakly to it.

PROOF OF PROPOSITION 7:

The first claim is immediate from the discussion in the text. To see the second claim, for each α denote the unique invariant probability measure of establishment sizes (see Proposition 5) 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 of the establishment size growth rate. We saw above that

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

So 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 for α_j , and the operator $T_{\alpha_j}^*$ maps the λ_{α_k} into distributions with thinner left tails. The cases for intermediate and high $\ln n$ are analogous.

REFERENCES

- ▶ **Albuquerque, Rui, and Hugo A. Hopenhayn.** 2004. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies*, 71(2): 285–315.
- ▶ **Audretsch, David B., Luuk Klomp, Enrico Santarelli, and A. Roy Thurik.** 2004. "Gibrat's Law: Are the Services Different?" *Review of Industrial Organization*, 24(3): 301–24.
- Axtell, Robert L.** 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293 (September 7): 1818–20.
- ▶ **Cabral, Luís M. B., and José Mata.** 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory." *American Economic Review*, 93(4): 1075–90.
- Caves, Richard E.** 1998. "Industrial Organization and New Findings on the Turnover and Mobility of Firms." *Journal of Economic Literature*, 36(4): 1947–82.
- Clementi, Gian Luca, and Hugo A. Hopenhayn.** 2006. "A Theory of Financing Constraints and Firm Dynamics." *Quarterly Journal of Economics*, 121(1): 229–65.
- Cooley, Thomas F., and Vincenzo Quadrini.** 2001. "Financial Markets and Firm Dynamics." *American Economic Review*, 91(5): 1286–1310.
- Davis, Steven J., and John Haltiwanger.** 1999. "Gross Job Flows." In *Handbook of Labor Economics*, Vol. 38, ed. Orley Ashenfelter and David Card, 2711–2805. Amsterdam: Elsevier Science, North-Holland.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson.** 1989a. "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector." *Journal of Labor Economics*, 7(1): 48–71.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson.** 1989b. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics*, 104(4): 671–98.
- Ericson, Richard, and Ariel Pakes.** 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1): 53–82.
- Evans, David S.** 1987a. "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy*, 95(4): 657–74.
- Evans, David S.** 1987b. "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries." *Journal of Industrial Economics*, 35(4): 567–81.
- Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–67.
- ▶ **Garicano, Luis, and Esteban Rossi-Hansberg.** 2004. "Inequality and the Organization of Knowledge." *American Economic Review*, 94(2): 197–202.
- Geroski, Paul A.** 1983. "The Empirical Analysis of Entry: A Survey." University of Southampton Working Paper 8318.
- Geroski, Paul A.** 1995. "What Do We Know about Entry?" *International Journal of Industrial Organization*, 13(4): 421–40.
- Geroski, Paul A., and Joachim Schwalbach,** eds. 1991. *Entry and Market Contestability*. Oxford: Blackwell.
- Gibrat, Robert.** 1931. *Les Inégalités Économiques*. Paris: Libraire du Recueil Sirey.
- Gorecki, Paul K.** 1976. "The Determinants of Entry by Domestic and Foreign Enterprises in Canadian Manufacturing." *Review of Economics and Statistics*, 58(4): 485–88.
- Griliches, Zvi, and Haim Regev.** 1995. "Firm Productivity in Israeli Industry 1979–1988." *Journal of Econometrics*, 65 (1): 175–203.
- Guner, Nezih, Gustavo Ventura, and Xu Yi.** 2005. "Macroeconomic Implications of Size-Dependent Policies." Unpublished.
- Hall, Bronwyn H.** 1987. "The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector." *Journal of Industrial Economics*, 35(4): 583–606.
- Hause, John C., and Gunnar Du Rietz.** 1984. "Entry, Industry Growth, and the Microdynamics of Industry Supply." *Journal of Political Economy*, 92(4): 733–57.
- Hopenhayn, Hugo A.** 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–50.
- Jones, Charles I.** 2005. "The Shape of Production Functions and the Direction of Technical Change." *Quarterly Journal of Economics*, 120(2): 517–49.
- Jovanovic, Boyan.** 1982. "Selection and the Evolution of Industry." *Econometrica*, 50(3): 649–70.
- Kambourov, Gueorgui, and Iourii Manovskii.** 2005. "Occupational Specificity of Human Capital." Unpublished.
- ▶ **Klette, Tor Jakob, and Samuel Kortum.** 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy*, 112(5): 986–1018.
- Lucas, Robert E., Jr.** 1978. "On the Size

- Distribution of Business Firms." *Bell Journal of Economics*, 9(2): 508–523.
- Luttmer, Erzo G. J.** Forthcoming. "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics*.
- MacDonald, James M.** 1986. "Entry and Exit on the Competitive Fringe." *Southern Economic Journal*, 52(3): 640–52.
- Mansfield, Edwin.** 1962. "Entry, Gibrat's Law, Innovation, and the Growth of Firms." *American Economic Review*, 52(5): 1023–51.
- Orr, Dale.** 1974. "The Determinants of Entry: A Study of the Canadian Manufacturing Industries." *Review of Economics and Statistics*, 56(1): 58–66.
- Restuccia, Diego, and Richard Rogerson.** 2004. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." Unpublished.
- **Rossi-Hansberg, Esteban, and Mark L. J. Wright.** 2007. "Urban Structure and Growth." *Review of Economic Studies*, 74(2): 597–624.
- Scherer, Frederic. M.** 1980. *Industrial Market Structure and Economic Performance*, 2nd ed. Chicago: Rand McNally College Pub. Co.
- Sleuwaegen, Leo, and Micheline Goedhuys.** 2002. "Growth of Firms in Developing Countries, Evidence from Côte d'Ivoire." *Journal of Development Economics*, 68(1): 117–35.
- Stokey, Nancy L., and Robert E. Lucas, Jr., with Edward C. Prescott.** 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.
- Sutton, John.** 1997. Gibrat's Legacy. *Journal of Economic Literature*, 35(1): 40–59.
- Tybout, James R.** 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38(1): 11–44.