Working Paper No. 518

Entry Costs Rise with Development

By

Albert Bollard
Peter J. Klenow
Huiyu Li

May 2014

Entry Costs Rise with Development

Albert Bollard

McKinsey Consulting

Peter J. Klenow[*]

Stanford University

Huiyu Li[†]

Stanford University

May 29, 2014

Preliminary

Abstract

Looking at manufacturing industries over time and across countries, the number of firms and establishments is closely tied to employment. Relative to employment, the number of businesses is at best weakly increasing in output per worker. In many models of firm dynamics, trade and growth, these facts imply that the cost of creating a new firm or plant increases sharply with productivity. This increase in entry costs can stem from rising cost of labor used in entry, as well as higher output costs of setting up a business to use more sophisticated technologies. How entry costs vary with development matters for welfare in many settings, such as in love of variety models and in span-of-control models. Our findings suggest that the welfare impact of productivity-enhancing policies is not significantly amplified through an increase in the number of firms or plants.

1. Introduction

As countries get richer, do they create more firms and establishments per worker, or just better ones? Policies that raise productivity can induce more firms or more output per firm, depending on the entry technology. Suppose for a moment that new businesses are created with a fixed amount of output. Then endogenous expansion in the number of firms can amplify the welfare gains from policies that boost productivity, say by increasing variety or reducing span-of-control costs. This multiplier effect through entry is analogous to the multiplier effect on output from physical capital accumulation in the neoclassical growth model. Now suppose instead that entry requires a fixed amount of labor. Then policies boosting productivity are not amplified through entry because entry costs rise with the price of labor.

Widely used models of firm dynamics, growth, and trade make different assumptions about entry costs.[1] Some models assume that entry costs are stable with development (e.g. a fixed output cost to invent a new product). Examples include Hopenhayn (1992), Romer (1994), Foster et al. (2008), and David (2013). Other models assume that entry costs rise with development, say because entry requires a fixed amount of labor and labor becomes more expensive with development. See, for example, Lucas (1978), Grossman and Helpman (1991), Melitz (2003), Klette and Kortum (2004), Luttmer (2007), and Acemoglu et al. (2013). Other studies do not take a stand but emphasize that the entry technology matters for the welfare impact of policies; see Rivera-Batiz and Romer (1991), Atkeson and Burstein (2010), Bhattacharya et al. (2012), and the survey by Costinot and Rodríguez-Clare (2013).

Existing evidence is limited on how entry costs change with development. This is why models are mixed or agnostic on the question. The evidence is mostly confined to estimates of the regulatory barriers to entry across coun-

---

[1]By "entry costs" we have in mind all non-production costs over a firm's life cycle. These include not only upfront innovation and setup costs but also overhead costs, R&D of incumbents, and fixed costs of exporting.

tries, to the exclusion of the technological costs of innovating and setting up operations. Djankov et al. (2002) document higher statutory costs of entry (relative to GDP per capita) in poor countries. Their pioneering effort spawned the influential Doing Business Surveys conducted by the World Bank. Luttmer (2007, 2010) reports that the U.S. firm size distribution appears stationary over time. He shows that making entry costs proportional to the level of development is necessary for the existence of a stationary firm size distribution in various growth models.

In this paper, we provide evidence on how the number of businesses varies with the level of labor productivity. We look over time in the U.S. (1982–2007 quinquennial Census data), China (1998-2007 Surveys of Industrial Production), and India (1989-2004 Annual Survey of Industries and Surveys of Unorganized Manufactures). We also look across countries (2006 and earlier UNIDO data). We look mostly at manufacturing industries, but present some evidence for all U.S. industries. As a corollary, we document how the number of enterprises varies with the number of workers. We argue that these simple empirical elasticities discipline the nature of entry costs in widely used models.

We find that the number of firms or establishments increases only modestly with the level of labor productivity both over time and across countries. The U.S. over-time evidence comes entirely from the U.S. Census, for which measurement should be most consistent. The cross-country results are more suspect because of differences in sample coverage (e.g., they only cover the formal sector). Still, the cross-country results are consistent with the over-time patterns, and would surely be reinforced by informal sector data.[2] The same patterns are seen across OECD countries alone, for which data are more comparable. These facts imply that revenue per enterprise increases sharply with the level of development. Enterprises evidently need more revenue to satisfy the

---

[2]There are many more informal establishments per worker in developing countries, so that the number of all establishments per worker may fall with development. See Hsieh and Klenow (2014) for evidence on informality in India and Mexico, and La Porta and Shleifer (2008) for evidence on other countries.

free entry condition in places with higher labor productivity. If higher revenue is associated with higher operating profits, then entry costs must be bigger for the zero profit condition to hold.[3]

Entry costs could rise with development simply because entry is labor-intensive and labor is expensive in productive economies. This would explain why entry moves closely with employment but not with output per worker. Entry costs could also rise with development because entrants set up more technologically sophisticated operations in more advanced economies. Our evidence is relevant for total entry costs, i.e. the sum of technological and regulatory barriers. If, as seen in the Doing Business surveys, regulatory entry costs increase modestly with development, then technological entry costs must be the dominant force pushing up entry costs with development.

Our findings have implications for modeling and policy. First, if the choice is between fixed entry costs in terms of labor or output, our evidence favors denominating entry costs in terms of labor. Second, our evidence is consistent with the assumption of rising innovation costs with technological progress, as is often assumed to obtain balanced growth in theory.[4] Third, productivity-enhancing policies have muted effects on entry.

The rest of the paper proceeds as follows. Section 2 describes a few simple models to illustrate why we care about the nature of entry costs. Section 3 presents evidence on how the number of businesses varies with development over time and across countries. Section 4 discusses the potential implications of empirical patterns for entry costs. Section 5 considers alternative interpretations. Section 6 concludes.

---

[3]We consider other possibilities, such as variable markups, firm exit rates, real interest rates, or firm growth rates. We will argue that these forces are too weak to explain the massive variation seen in revenue per business.

[4]E.g. Romer (1990), Aghion and Howitt (1992), Kortum (1997), and chapters 13 and 14 of Acemoglu (2011).

2. Simple Motivating Models

Here, we use several models to illustrate that entry costs matter for welfare in many different settings.

2.1. Love-of-Variety

Consider a static, closed economy version of the Melitz (2003) model. The economy has a representative household endowed with L units of labor. Consumption per capita, which is equal to the real wage w, is a measure of welfare in the economy.

Consumption goods are produced by a perfectly competitive sector that uses intermediate goods as inputs and a CES production technology. Profit maximization yields a downward sloping demand curve for each intermediate good.

The intermediate goods sector is monopolistically competitive. Without loss of generality, we assume all firms in this sector have the same production function, which is linear in labor inputs with technology level A_y.[5] Each intermediate goods firm takes demand for its product as given and chooses its output or price to maximize its profit. This yields the familiar relationship between the wage bill, revenue, and profit in each firm

$$wl \ = \ \frac{\sigma - 1}{\sigma} py \ = \ (\sigma - 1)\pi \ \propto \ w^{1-\sigma} Y A_y^{\sigma-1} \tag{1}$$

where Y is aggregate output and $\sigma > 1$ is the elasticity of substitution between varieties. Let L_y be the total amount of labor devoted to producing intermediate goods. By symmetry of the intermediate goods production function

$$Y = A_y L_y M^{\frac{1}{\sigma-1}}. \tag{2}$$

One unit of an entry good is required to create a variety, i.e., set up an inter-

---

[5]We could generalize to allow post-entry heterogeneity in firm technology and define $A_y := (\mathbb{E} A_y^{\sigma-1})^{\frac{1}{\sigma-1}}$.

mediate goods firm. We generalize the production technology of the entry good
in Melitz (2003) to allow final goods to be an input into creating a new variety.
In particular, we follow Atkeson and Burstein (2010) in assuming that the entry
technology has the Cobb-Douglas form

$$M = A_e Y_e^\lambda L_e^{1-\lambda} \tag{3}$$

where M is the number of varieties created, and L_e and Y_e are the amount of
labor and final output, respectively, used in creating varieties.

Perfect competition in the CRS sector producing entry goods implies that
the cost of creating a variety in terms of consumption goods is

$$p_e \propto \frac{w^{1-\lambda}}{A_e}. \tag{4}$$

And the labor share of revenue in entry goods production is

$$wL_e = (1 - \lambda)p_e M. \tag{5}$$

Free entry, with positive entry in equilibrium, implies

$$\pi = p_e \tag{6}$$

which equates profit per variety to the entry cost.

Thus, the one-shot equilibrium given (L, A_y, A_e) consists of prices (w, p_e) and
allocations (C, M, Y, L_e, L_y) such that $C = wL$, (1), (2), (3), (4), and (6) hold, and
the following labor and goods market clearing conditions are satisfied:

$$L = L_y + L_e, \quad Y = C + Y_e. \tag{7}$$

We now consider how the welfare impact of a change in A_y depends on the

entry technology. In equilibrium, welfare (the real wage) is

$$w = A_y M^{\frac{1}{\sigma-1}} \tag{8}$$

so

$$\frac{\partial \ln w}{\partial \ln A_y} = 1 + \frac{1}{\sigma - 1} \frac{\partial \ln M}{\partial \ln A_y}.$$

An increase in A_y not only raises welfare directly, but also has the potential to improve welfare indirectly through variety expansion.

One can show that equilibrium variety satisfies

$$M \propto \frac{wL}{p_e} \tag{9}$$

so that the number of varieties depends on the value of labor relative to the entry cost. Combining this with equation (6) relating the real wage to p_e, we get

$$\frac{\partial \ln M}{\partial \ln A_y} = \lambda \frac{\partial \ln w}{\partial \ln A_y}$$

That is, the elasticity of variety with respect to A_y is larger when the share of output used in producing varieties (λ) is bigger. Higher A_y means more output, and some of this output is devoted to producing more varieties if the final good is used in entry ($\lambda > 0$). Repeated substitution will show that the compounding impact of A_y on welfare is

$$\frac{\partial \ln w}{\partial \ln A_y} = 1 + \frac{\lambda}{\sigma - 1 - \lambda}$$

with the second term capturing the effect of variety expansion. A higher output share (λ) means more amplification.

The amplification of an increase in productivity depends on σ, the degree of substitutability of intermediate goods, because varieties are more valuable when substitutability is low. To illustrate the potential importance of variety

expansion, consider the Broda and Weinstein (2006) estimates of $\sigma \approx 4$ at the 3-digit to 4-digit level. For this value of σ, the amplification can range from 50% when $\lambda = 1$ to 0% when $\lambda = 0$. Thus, for a plausible value of σ, the nature of entry costs matters materially for the welfare impact of changes in production technology A_y.

The entry technology also influences the welfare impact of policies affecting the level of the population or allocative efficiency. As in Melitz (2003), increasing the population is like an extreme trade liberalization going from autarky to frictionless trade between countries. In this case, the overall welfare effect is

$$\frac{\partial \ln w}{\partial \ln L} = \frac{1}{\sigma - 1} \left(1 + \frac{\lambda}{\sigma - 1 - \lambda} \right)$$

Again, at $\sigma = 4$ the amplification through variety expansion is 50% when $\lambda = 1$ and 0% when $\lambda = 0$.

Now, it is plausible that different production technologies have intrinsically different setup costs. Suppose that, in the previous model, the entry technology parameter in (3) is related to the production technology by

$$\ln A_e = -\mu \ln A_y + \epsilon$$

where ϵ is a component unrelated to A_y and μ captures how fast entry costs rise with production technology (for a given cost of labor). In this case we still have

$$\frac{\partial \ln w}{\partial \ln A_y} = 1 + \frac{1}{\sigma - 1} \frac{\partial \ln M}{\partial \ln A_y}.$$

But now

$$\frac{\partial \ln M}{\partial \ln A_y} = \lambda \frac{\partial \ln w}{\partial \ln A_y} + \frac{\partial \ln A_e}{\partial \ln A_y} = \lambda \frac{\partial \ln w}{\partial \ln A_y} - \mu.$$

The welfare impact of a change in the production technology becomes

$$\frac{\partial \ln w}{\partial \ln A_y} = 1 + \frac{\lambda - \mu}{\sigma - 1 - \lambda}.$$

Thus, when entry costs rise with productivity, either through higher labor costs (small λ) or higher costs of setting up more sophisticated businesses (large positive μ), the impact of A_y on variety and welfare is dampened.

2.2. Span-of-Control

The entry technology matters for welfare even in a Lucas span-of-control model in which there is no love-of-variety. Consider the environment

$$Y = \sum_{i=1}^{M} Y_i$$

$$Y_i = A_y L_i^{\gamma}$$

$$M = A_y^{-\mu} L_e^{1-\lambda} Y_e^{\lambda}$$

The first equation says aggregate output is the simple sum of firm output levels. The second equation specifies diminishing returns to production labor for each firm ($\gamma < 1$). The third equation is the technology for entry. Whereas Lucas (1978) specified overhead costs due to a single manager's time, we allow for the possibility that overhead involves goods as well as labor. Bloom et al. (2013) for example, argue that overhead costs include some information technology equipment. Variable profits are then

$$\pi_i = (1 - \gamma)Y_i = A_y^{\frac{1}{1-\gamma}} \left(\frac{\gamma}{w}\right)^{\frac{\gamma}{1-\gamma}}.$$

As in the love-of-variety model, free entry implies

$$\pi_i = p_e \propto A_y^{\mu} w^{1-\lambda}.$$

In general equilibrium

$$\ln w = \frac{\sigma - 1 - \mu}{\sigma - 1 - \lambda} \ln A_y + constant$$

$$\ln \frac{Y}{M} = \left(1 - \frac{(\sigma-1)(\lambda-\mu)}{\sigma-1-\mu}\right) \ln \frac{Y}{L_y} + constant$$

The welfare impact of a change in A_y here is the same as in the love-of-variety model when $1 - \gamma = \frac{1}{\sigma-1}$. If better production technology boosts entry, then production labor is spread more thinly across firms, limiting scale diseconomies. Thus entry can amplify the welfare impact of better technology, just as in the love-of-variety model. Unlike in the love-of-variety model, however, changes in L do not affect welfare. A bigger population increases the number of firms proportionately, but leaves aggregate productivity unchanged.

To recap, the entry technology (parameterized by λ and μ) matters for welfare analysis in the span-of-control model.

2.3. Growth with Quality Ladders and Expanding Varieties

Consider a sequence of one-shot-economies, as in the love-of-variety model, with the following modifications: 1) knowledge spillovers from period $t-1$ to t; and 2) each entrant chooses its quality (process efficiency) A_t and the number of varieties v_t it will produce.

In each period t, the past pool of knowledge A_{t-1} improves the current entry technology:

$$p_t^e \propto e^{\mu \frac{A_t}{A_{t-1}}} f(v_t, A_t) w_t^{1-\lambda} =: \frac{w_t^{1-\lambda}}{A_t^e}$$

where p_t^e is the entry cost for a firm. An entering firm chooses its quality level A_t and the number of varieties to produce v_t. Profit maximization and free entry

imply that

$$\frac{\partial \ln \pi_t(A_t, v_t)}{\partial \ln A_t} = \frac{\partial \ln p_t^e}{\partial \ln A_t}$$

and

$$\frac{\partial \ln \pi_t(A_t, v_t)}{\partial \ln v_t} = \frac{\partial \ln p_t^e}{\partial \ln v_t}.$$

Variable profits are $\pi_t(A_t, v_t) = \pi_t A_t^{\sigma-1} v_t$, so the firm's optimal choice of A_t satisfies

$$\sigma - 1 = \mu \frac{A_t}{A_{t-1}} + \frac{f_A(v_t, A_t)}{f(v_t, A_t)} A_t$$

and its optimal choice of v_t is given by

$$1 = \frac{f_v(v_t, A_t)}{f(v_t, A_t)} v_t.$$

Assume

$$f(v, A) = e^{\frac{v^\rho}{A}}, \rho > 1$$

so that the marginal cost of producing an additional variety in a firm is increasing in the number of varieties produced in the firm, and choosing a higher technology level lowers the overall cost of producing varieties in a firm.[6] This particular functional form implies that the growth rate of quality between $t-1$ and t is

$$g_t^A := \ln \frac{A_t}{A_{t-1}} = \ln \frac{\sigma - 1 + \frac{1}{\rho}}{\mu}$$

and the number of varieties per firm grows at

$$g_t^v := \ln \frac{v_t}{v_{t-1}} = \frac{1}{\rho} g_t^A$$

The equilibrium number of firms per worker is

$$\ln \frac{N_t}{L_t} = \lambda \ln \frac{Y_t}{L_t} - \ln f(v_t, A_t) + constant$$

---

[6]We want to allow increases in quality over time to facilitate growing variety per firm, as there is evidence of such trends in U.S. data. See Bernard et al. (2010) and Broda and Weinstein (2010).

where N_t is the number of firms. The number of varieties produced in the economy is $M_t := N_t v_t$. The real wage and hence welfare in this economy is

$$\ln w_t = \frac{\sigma - 1}{\sigma - 1 - \lambda} \left(\ln A_t + \frac{\ln L_t v_t - \ln f(v_t, A_t)}{\sigma - 1} \right) + constant$$

and the growth rate of the real wage is

$$g_t^w := \frac{g^L + g^A(\sigma - 1) + g^v}{\sigma - 1 - \lambda}.$$

Similar to the static love-of-variety model, a higher λ implies a larger welfare effect of changes in the level and growth rate of A_t and L_t. This model illustrates that amplification through entry can occur in an endogenous growth model with rising quality, expanding variety, and population growth – and in which firms produce multiple varieties. In particular, amplification is from variety expansion through an increase in the number of firms, whether or not there are multiple or even growing varieties per firm.

3.   Empirical Patterns

Having motivated why we care about how entry costs vary with development, we now attempt to provide evidence relevant to the question. In this section, we present results for OLS regressions of the form:

$$\ln \frac{M_i}{L_i} = \beta_0 + \beta_1 \ln \frac{Y_i}{L_i} + \beta_2 Firm\_Dummy_i + \beta_3 Industry\_Dummy_i + \epsilon_i \quad (10)$$

where M is the number of establishments or firms (depending on data availability), Y is value added, and L is the number of workers. When the sample is a mix of firm and establishment data, we put in a $Firm\_Dummy$ for countries with firm data to take into account that firms are typically bigger than establishments. We also control for industry fixed effects.[7] Subscript i is a year over

---

[7] β_3 and $Industry\_Dummy_i$ are vectors.

time in the U.S., China or India in the first subsection below. In the second subsection, i is a country or a country-industry pair in 2006. We focus on manufacturing almost exclusively because of data availability, and because free entry is more likely there than in (sometimes heavily) regulated sectors such as mining, utilities, or finance.

Regression equation (10) is motivated by the entry goods production function (3). The Cobb-Douglas technology for entry implies that the number of establishments relative to entry labor is proportional to the real wage raised to the power λ, the goods share in the entry production function. Meanwhile, CES aggregate of goods implies a constant markup for producers, and hence revenue per worker proportional to the real wage. As a result, the real wage w and the quantity of entry per entry labor are related to output, employment, and the number of firms (Y, L and M) by

$$\frac{M}{L} \propto \frac{M}{L_e} = A_e \left(\frac{Y_e}{L_e} \right)^{\lambda} \propto A_e w^{\lambda} \tag{11}$$

$$w = \frac{\sigma - 1}{\sigma} \frac{Y}{L} \frac{L}{L_y} = \frac{1 - \lambda + \sigma - 1}{\sigma} \frac{Y}{L}. \tag{12}$$

Thus regressing M/L on Y/L, both in logs, should reveal the importance of goods in the entry technology (λ) in this simple model. The residual in this relationship should be the entry technology shifter A_e. For now, imagine that this residual is orthogonal to output per worker, so that OLS will yield consistent estimates of λ.[8] Figure 1 displays the model's prediction for two polar cases of λ, and previews what we will report from the data in the rest of this section.

3.1. Over time in the U.S., China and India

We start by looking at U.S. Census data, which covers all establishments with employees. Manufacturing-wide data is publicly available every five years from

---

[8]In the next section we will take into account the endogeneity of value added per worker to the entry technology parameter A_e, and deploy IV in pursuit of consistent estimates. For this section, we present the OLS patterns for simplicity and transparency.
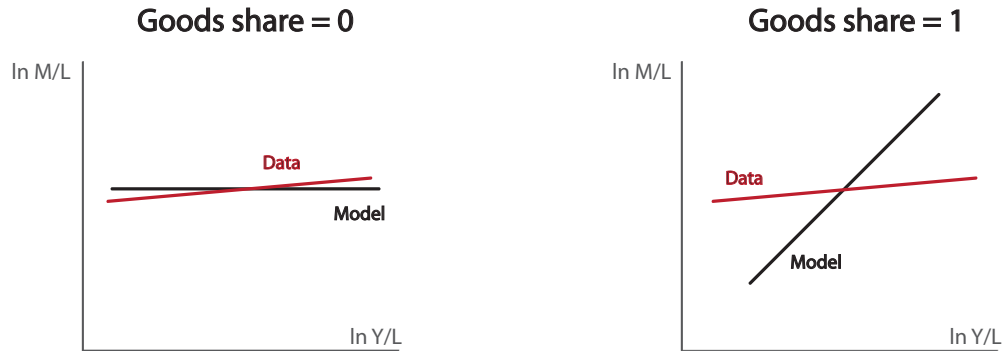
Figure 1: Model vs. data under $\lambda = 0$ (left panel) vs. $\lambda = 1$ (right panel).

1982 to 2007.

The first plot in Figure 2 shows the six Census years (1982, 1987, ..., 2007) for all of manufacturing. Value added is deflated by the U.S. Bureau of Economic Analysis manufacturing value added deflator. As shown, number of establishments per worker increases modestly with value added per worker. As real labor productivity has grown in the U.S., plants per worker has grown with an elasticity of around 0.2.

The Annual Survey of Manufacturing has annual data for value added and employment from 1982-2009 but does not have establishment counts. On the other hand, Business Dynamics Studies provides data on the number of establishments and employment but does not have data on value added. In the second and third plot of Figure 2, we construct establishments per worker and value added per worker using different combinations of these two datasets. The regression results echo the aforementioned finding based on the Census data.

Figures 3 and 4 show the variation over time in establishments per worker vs. revenue per worker in Chinese and Indian manufacturing. Our estimates of value added, establishments and employment for China are from the Surveys of Industrial Production, and are similar to those tabulated independently by Brandt et al (2012). For India, we combine plant level data from the Annual Survey of Industries and the Survey of Unorganized Manufactures for 1989, 1994,
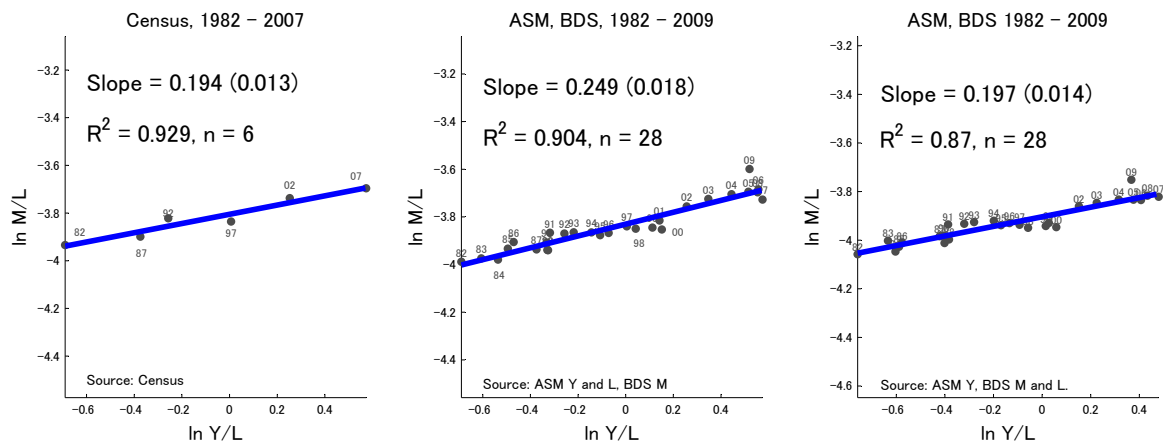
Figure 2: Establishments vs. output per worker, U.S. manufacturing over time

1999, 2004 and 2009. The Indian data includes the informal sector that tends to be excluded from plant level data on developing countries. Our findings for China and India bracket our results for the U.S. In China the number of establishments rises more steeply with average labor productivity, with an elasticity of 0.4. In India the elasticity is basically zero. Taken together, the evidence for all three countries is that establishments per worker increase with labor productivity with an elasticity of 0.2 plus or minus 0.2.

3.2. Cross-country evidence

Our cross-country dataset is the United Nations Industrial Development Organization (UNIDO) Industrial Statistics Database at the 4-digit level of ISIC Code Revision 3 (INDSTAT4 2012 ISIC Rev.3) for manufacturing.[9] We use the 2006 data because it has the largest number of countries. The data is from the OECD for member nations, and from national statistical offices for non-OECD countries. We use series on the number of employees, number of enterprises

---

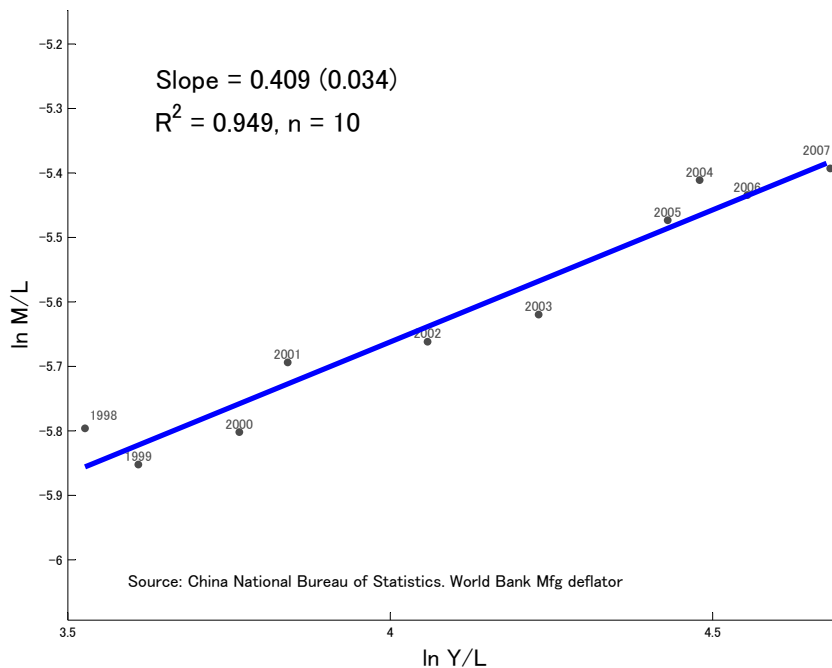[9]There are 127 4-digit ISICs with data for at least one country in 2006.

Figure 3: Establishments vs. value added per worker, Chinese manufacturing over time
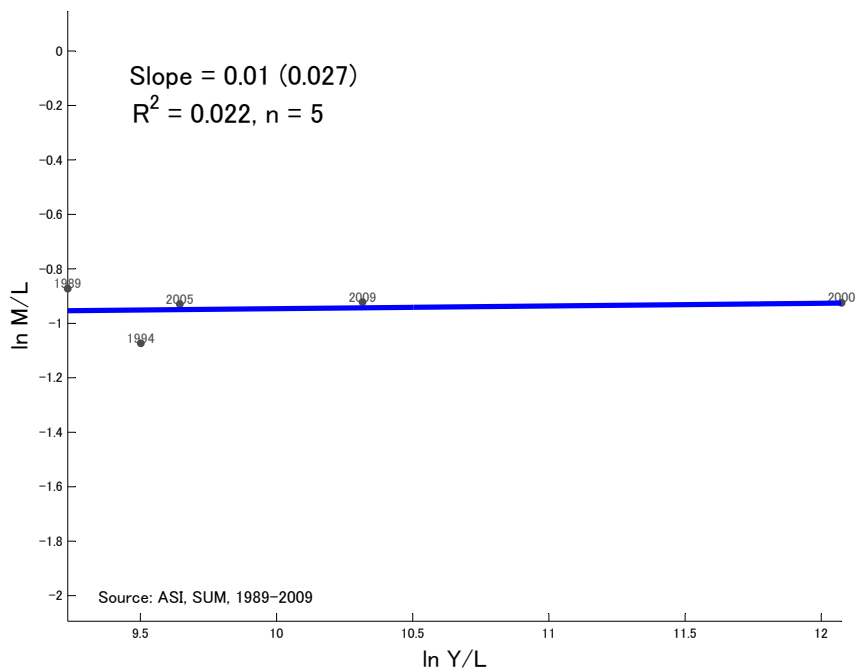

Figure 4: Establishments vs. value added per worker, India manufacturing over time

(firms), number of establishments, and value added. Not all countries report the number of establishments so we use the number of enterprises when establishment counts are not available. We keep only countries with data on the number of employees, value added, and the number of enterprises or establishments. This selection leaves us with 72 countries, of which 33 have establishment data. To compare value added across countries, we use the U.S. dollar values from UNIDO, which are converted from national currencies at IMF market/official exchange rates.[10]

Figures 5 to 7 display the cross-country data. Figure 5 shows that establishments (or firms) per worker increases in output per worker with an elasticity of around 0.2. Figure 6 is the counterpart of Figure 5 that controls for industry fixed effects at the ISIC 4-digit level. The slope remains modest at around 0.1 when looking across countries within 4-digit industries, so the aggregate cross-country relationship is not being driven by industry composition across countries.[11] To further demonstrate robustness across industries, we ran the $\ln M/L$ on $\ln Y/L$ regression allowing for industry-specific coefficients on $\ln Y/L$ for industries with at least 50 countries. Figure 7 plots the distribution of the coefficients. The interquartile range of the coefficients on $\ln Y/L$ is 0.786 to

---

[10]Deviations from PPP are much more important for nontradables than for tradables, and within manufacturing may owe to nontradable local distribution rather differences in manufacturing prices. See Hsieh and Klenow (2007).

[11]We use value-added shares because a natural extension of the static model in Section 2 has the property that the coefficient from the aggregate regression is equal to the value-added weighted average of the industry specific coefficients. More precisely, if we assume production technology

$$Y = \prod_s Y_s^{\theta_s}, \quad M_s = A_{e,s} L_{R,s}^{1-\lambda_s} Y_{R,s}^{\lambda_s}$$

then at the equilibrium

$$\ln \frac{M_s}{L_s} = constant_s + \lambda_s \ln \frac{P_s Y_s}{L_s} + \ln A_{e,s}$$

and

$$\ln \frac{M}{L} = constant + \left(\sum_s \frac{P_s Y_s}{Y} \lambda_s \right) \ln \frac{Y}{L} + \sum_s \frac{P_s Y_s}{Y} \ln A_{e,s}, \quad \text{where } M := \sum_s M_s.$$
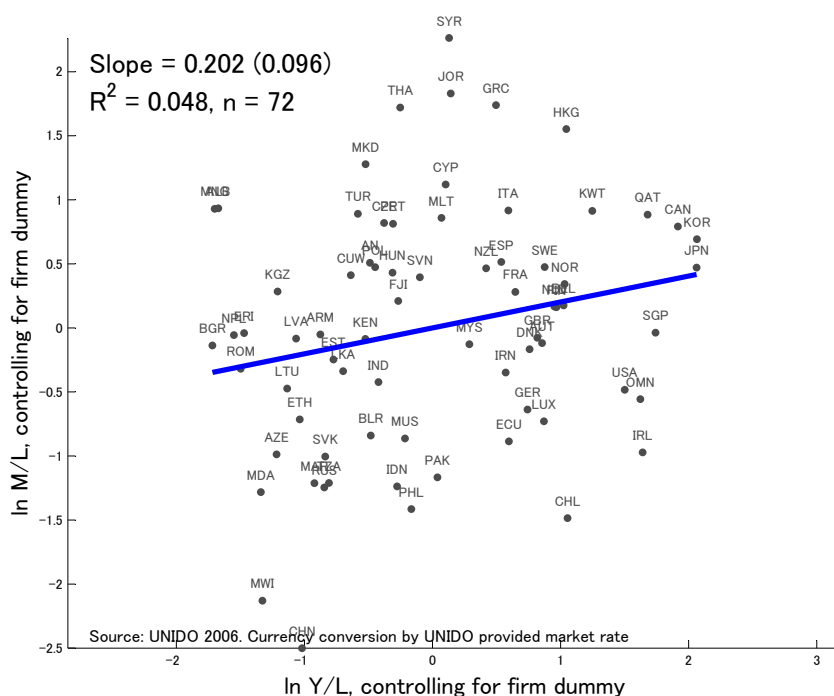
Figure 5: Cross-country variation in number of establishments (or firms) per worker with value added per worker, manufacturing
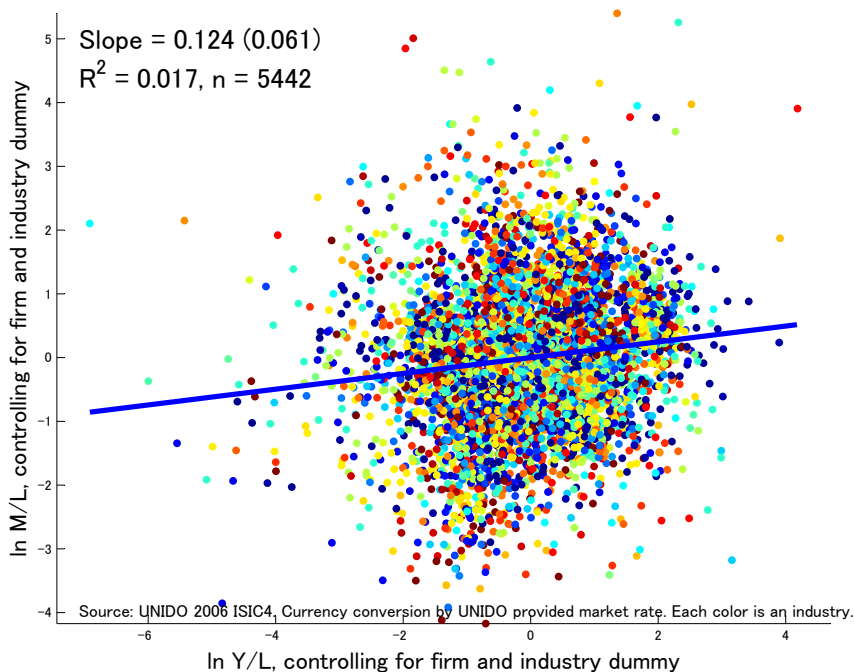

Figure 6: Cross-country variation in number of establishments (or firms) per worker with value added per worker, conditional on manufacturing industry (ISIC4)
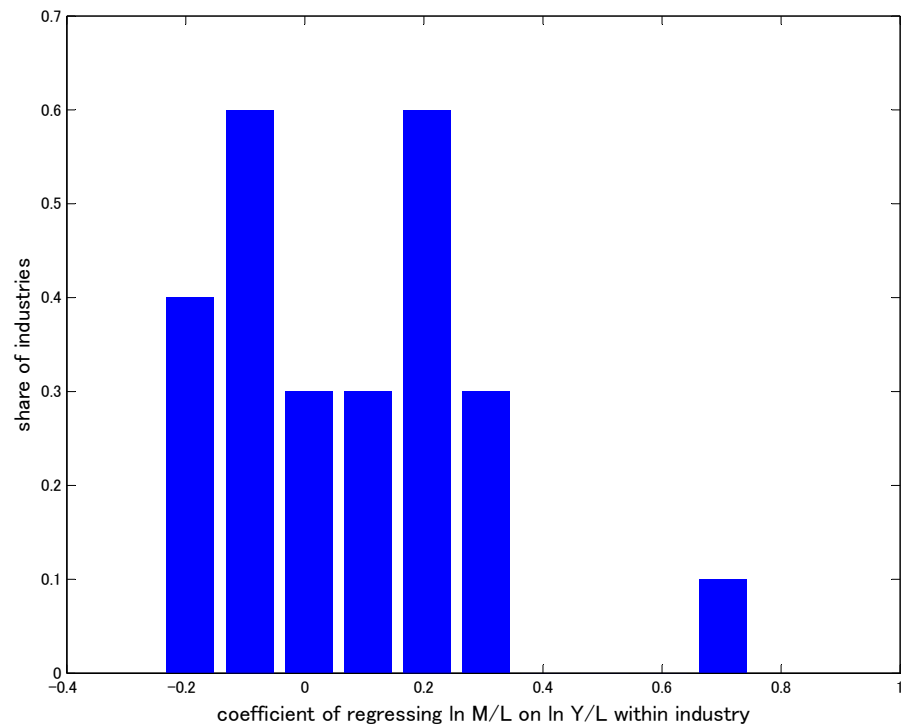
Figure 7: Distribution of regression coefficients of number of establishments (or firms) per worker on value added per worker across countries within manufacturing industry (ISIC4)

1.097, and the interquartile range of the standard error on the slope coefficient is 0.064 to 0.165. Thus the modest cross-country relation between number of establishments per worker and output per worker is a robust property across industries.[12]

Table 1: Cross country OLS regression of number of firms/establishments per worker on value added per worker

Dep Var: $\ln M/L$	(1)	(2)	(3)	(4)	(5)
Countries:	All	All	No size threshold	OECD	All
Firm or estab:	Both	Both	Both	Both	Esta
Firm dummies:	Yes	Yes	Yes	Yes	-
Industry f.e.	-	ISIC 4	ISIC 4	ISIC 4	ISIC4
$\ln Y/L$	0.202	0.124	0.018	-0.107	0.085
	(0.096)	(0.061)	(0.066)	(0.146)	(0.074)
R^2	0.048	0.017	0	0.007	0.01
Sample size	72	5442	3870	2705	2024

Standard errors are clustered by country for the ISIC-4 regressions.

Table 1 provides results from regressing plants (or firms) per worker on value added per worker – both in natural logs. The first two columns are analogues to the figures. The next three columns provide additional specifications. The third column includes only the 50 countries for which the sampled firms have no size threshold. The fourth column uses only data from OECD nations, for which

---

[12]The results are robust to using 3-digit controls instead, which increase the number of plants per industry-country.

measurement is more consistent and for which the missing informal manufacturing sector is presumably much smaller.[13] The fifth and final column restricts the countries to those with establishment counts (rather than pooling firm and establishment data). In all specifications, the slope is 0.2 or lower.

4. The entry cost explanation

4.1. Entry costs

If the zero profit condition holds for entrants as a whole and markups are fixed, then average firm value added can proxy for entry costs. This is equivalent to the number of businesses expanding almost in proportion to the population. Thus, one interpretation of our regression results is that entry costs rise sharply with development. For the simplified Melitz model in Section 2, assuming $A_e = A_y^{-\mu} e^\epsilon$ and using $\ln Y/M$ to proxy for entry costs, the following relationship holds between observables:

$$\ln M = constant + \frac{(\lambda - \mu)(\sigma - 1)}{\sigma - 1 - \mu} \ln \frac{Y}{L} + \frac{\sigma - 1}{\sigma - 1 - \mu} \ln L + \frac{\sigma - 1}{\sigma - 1 - \mu}\epsilon. \quad (13)$$

When $\mu = 0$, this equation is the regression equation we ran in Section 3 but with $\ln L$ on the RHS.

Note, however, that Y/L is endogenous to ϵ — years and countries with higher idiosyncratic entry costs should have fewer businesses and hence lower labor productivity. As a result, the coefficients we obtained in the previous section's OLS regressions using (13) should not generate consistent estimates even if this simple model was perfectly describing the data. One can deal with this endogeneity issue if instruments are available.

To illustrate the potential magnitude of OLS endogeneity bias, suppose that

---

[13]For countries outside the OECD, the missing informal manufacturing sector can be quite large. See Tybout (2000) and Hsieh and Klenow (2014).

$\epsilon \perp \ln A_y$ and $\epsilon \perp \ln L$. I.e., suppose that idiosyncratic entry costs are orthogonal to both the production technology and population in a country. Tables 2 and 3 display the results of GMM estimation using these moment restrictions on U.S. time series from 1982–2009 and UNIDO countries in 2006, respectively. The first row shows that $\lambda \geq 0$ is binding.[14] The implication is that entry requires labor but does not become more difficult with better production technology (μ is negative in Table 2 and positive but small in Table 3).

Table 2: Estimating μ and λ using US time series data

Identifying restrictions	$\hat{\lambda}$	$\hat{\mu}$	$\frac{\hat{\lambda}-\hat{\mu}}{\sigma-1-\lambda}, \sigma = 4$ (amplification)
$\epsilon \perp \ln A_y, \epsilon \perp \ln L$	0	-0.181 (0.822)	0.060
$\epsilon \perp \ln A_y, \epsilon \perp \ln L, \mu = 0$	0.181 (0.012)	0	0.064
$\epsilon \perp \ln A_y, \mu = 0$	0.198 (0.014)	0	0.071
$\epsilon \perp \ln A_y, \epsilon \perp \ln L, \lambda = 1$	1	0.808 (0.015)	0.096
$\epsilon \perp \ln A_y, \lambda = 1$	1	0.786 (0.017)	0.107

Note: Estimates of λ without standard errors are cases when the estimate is at its lower bound of 0.

When we impose $\mu = 0$, in the second row, we find that λ is around 0.2 (U.S. time series) or continues to be at the zero lower bound (cross-country). With a single parameter to estimate, we can also relax one of the moment restrictions.

[14]We impose $\lambda \in [0, 1]$ so that the share of goods and labor in entry are nonnegative.

Table 3: Estimating μ and λ using cross country data

Identifying restrictions	$\hat{\lambda}$	$\hat{\mu}$	$\frac{\hat{\lambda}-\hat{\mu}}{\sigma-1-\lambda}, \sigma = 4$ (amplification)
$\epsilon \perp \ln A_y, \epsilon \perp \ln L$	0	0.041 (0.243)	-0.013
$\epsilon \perp \ln A_y, \epsilon \perp \ln L, \mu = 0$	0	0	0
$\epsilon \perp \ln A_y, \mu = 0$	0.016 (0.144)	0	0.005
$\epsilon \perp \ln A_y, \epsilon \perp \ln L, \lambda = 1$	1	0.706 (0.125)	0.147
$\epsilon \perp \ln A_y, \lambda = 1$	1	0.779 (0.117)	0.111

Note: Estimates of λ without standard errors are cases when the estimate is at its lower bound of 0.

When we no longer assume employment is orthogonal to idiosyncratic entry costs, λ remains close to 0.2 (U.S. time series) or close to zero (cross-country). Just like the OLS regressions, these IV estimates are consistent with entry costs being labor-intensive.

We can alternatively impose $\lambda = 1$ so that entry requires goods not labor, and see if this forces better technology to be more costly to set up ($\mu \gg 0$). Indeed, we estimate a μ between 0.7 and 0.8, depending on whether we impose both moment restrictions or only one, and whether we are looking at U.S. time series or UNIDO cross-country data. The assumption that idiosyncratic entry costs are orthogonal to the production technology ($\epsilon \perp \ln A_y$) is probably the most defensible, as $\mu > 0$ should incorporate the systematic component.
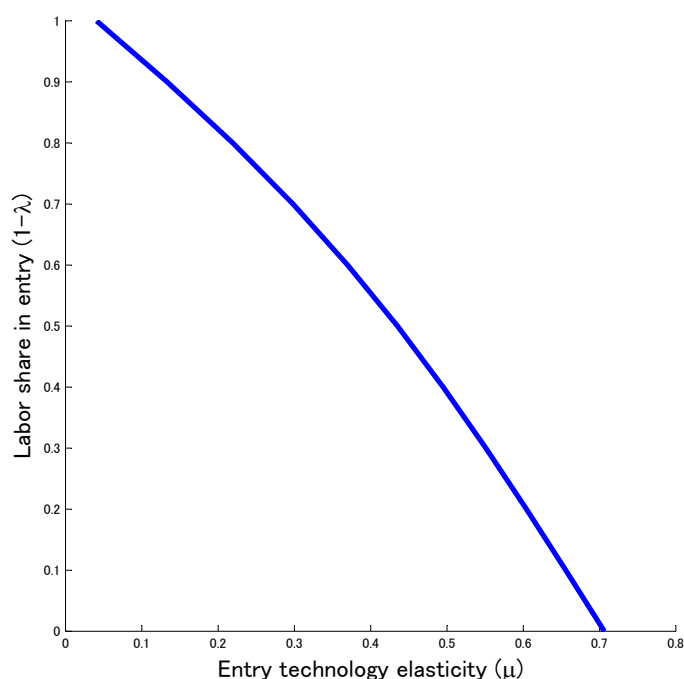
Figure 8: Combinations of $(1 - \lambda, \mu)$ from GMM estimation.

In all cases, the estimates of λ and μ imply modest entry expansion in re-sponse to better production technology – on the order of 0-15%, compared to the 50% one would have obtained with $\lambda = 1$ and $\mu = 0$. Figure 8 reinforces the point by showing the tradeoff between the labor share in entry costs $(1 - \lambda)$ and μ for the cross-country data. Entry costs must rise with the production tech-nology through some combination of more expensive labor $(\lambda < 1)$ or requiring more goods $(\mu > 0)$.

4.2.   Relation to Doing Business entry costs

Our inference that entry costs rise with development may seem at odds with a large literature on higher government-imposed entry barriers in poorer coun-tries. But one can think of total entry costs as the sum of a government-imposed

part and a technological part. Figure 9 plots our estimate of overall entry costs vs. the goverment-imposed costs in the World Bank Doing Business Survey, both against average value added per worker.[15] First, while the Doing Business costs decrease relative to income per capita, they increase in absolute terms with development. Second, the increase is much less steep than what we infer for overall entry costs, suggesting that the part of entry costs not captured by the Doing Business Survey increases faster with development than the costs measured by the Doing Business Survey. Third, overall entry costs seem to be much larger than the Doing Business Survey costs, suggesting that technological factors may constitute a larger part of entry costs than legal barriers to entry.

5. Competing explanations

5.1. Markups

Suppose entry costs are the same across countries but the price/cost markup varies. In the model of Section 2.1,

$$\left(\frac{Y}{M}\right)_{rich} - \left(\frac{Y}{M}\right)_{poor} = \left(\sigma_{rich} - \sigma_{poor}\right) p_e.$$

Richer economies could have larger firms because of lower markups rather than higher entry costs. Bresnahan and Reiss (1991) famously found that firms are bigger in more densely populated areas in the U.S., and argued it was because markups dropped with the number of competitors. Melitz and Ottaviano (2008) and Edmond et al. (2012) make a similar argument in trade models.

---

[15]We proxy entry costs by the present discounted value of profits. More precisely, using the model we layout in the next section, we calculate total entry cost as $\frac{Y}{M} \frac{1}{\sigma\delta} \frac{1-(1+g)^{\sigma-1}(1-\delta)}{1-\frac{(1+g)^{\sigma-1}(1-\delta)}{1+r}}$, with real interest rate $r = 0.05$, exit rate $\delta = 0.1$, and post-entry growth rate relative to entrants $g = 0.02$. We use the 2006 UNIDO data for average value added per firm. For government-imposed entry costs, we use the 2006 Doing Business Survey Starting a Business Costs. We calculate the government-imposed costs by multiplying the "Cost (% of income per capita)" series from the Doing Business Survey with $\frac{Y}{L}$ from UNIDO.
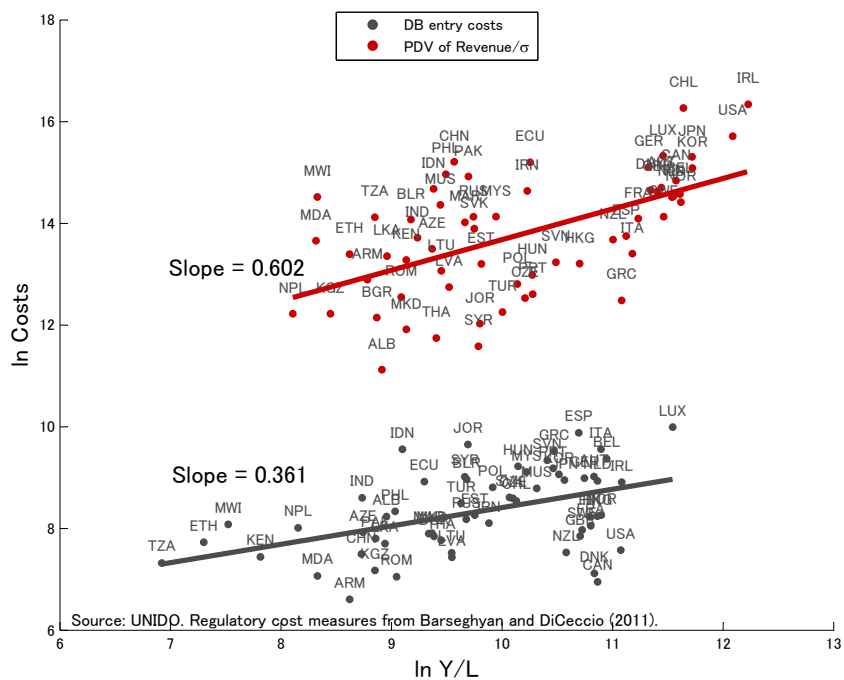
Figure 9: Overall entry costs vs. Doing Business costs.

Table 4: Required markup variation

Country	Required markup when
	$\lambda = 1, \mu = 0$ and U.S. markup is 1.33
India	∞
China	∞
France	1.97
Germany	1.84
UK	1.75
Japan	1.19

It is unlikely, however, that markup variation is large enough to explain our regression results. Under the assumption that entry costs are uniform across countries, we have

$$\left(\frac{Y}{M} \right)_{US} \Big/ \left(\frac{Y}{M} \right)_{j} = \sigma_{US} / \sigma_{j} .$$

Using $\sigma_{US} = 4$ from Broda and Weinstein (2006), we can compute the markup $\sigma/(\sigma - 1)$ that is required to explain average value added per firm in each country. Table 4 displays the computed markup for selected countries. Markups would have to be infinite in places like China and India, where value added per firm is much smaller than the U.S. If entry costs were the same in China and India as in the U.S., then all of output would have to be devoted to covering entry costs for firms to make zero profits there. This would leave no resources for consumption and investment. De Loecker et al. (2012) estimate much more modest markups in India – a median of 1.04 and a mean of 1.67. Even in other rich countries, such as France, Germany and U.K., the Table shows that markups would have to be more than double the level in the U.S.

5.2.   Discount rate, growth rate and exit rate

We extend the simple model in Section 2 to investigate whether differences in firm life-cycles across countries and over time within a country can explain our empirical findings. Suppose that entrants enter with productivity A_0, incumbents' productivity grows at rate g after entry, incumbents exit at exogenous rate δ, and firms discount future profits at rate r. Also assume that g is small enough relative to δ so that the economy is in a steady state with a stationary size distribution. The free entry condition in this model then equalizes the entry cost with the expected sum of discounted profits:

$$p_e = \Pi A_0^{\sigma-1} \sum_{a=0}^{\infty} \left(\frac{(1+g)^{\sigma-1}(1-\delta)}{1+r} \right)^a.$$

Due to stationarity, average revenue across firms at a point in time equals average revenue over a firm's lifetime, which is the expected sum of *undiscounted* revenue divided by the firm's expected lifespan $1/\delta$. That is,

$$\frac{Y}{M} = \delta\sigma\Pi A_0^{\sigma-1} \sum_{a=0}^{\infty} \left((1+g)^{\sigma-1}(1-\delta) \right)^a.$$

Hence, average revenue and the entry cost are related by

$$\frac{Y}{M} = \sigma\delta \frac{1 - \frac{(1+g)^{\sigma-1}(1-\delta)}{1+r}}{1 - (1+g)^{\sigma-1}(1-\delta)} p_e.$$

Average revenue increases relative to the entry cost when there is heavier discounting (higher r) or when revenue is back-loaded (higher g). When the exit rate increases, average revenue can either fall or rise. This nonmonotonicity arises because a higher exit rate works like higher discount rate that reduces the present discounted value of profit, but also puts less weight on future profits and reduces the discrepancy between revenue per firm and revenue over a firm's lifetime that occurs due to discounting. For example, when $r = 0$, average revenue relative to entry costs increases monotonically in the exit rate.

The above relationship between average revenue and entry costs suggests that discount rates, post-entry growth rates, and exit rates could potentially explain why average revenue increases with development across countries or over time within a country even when the entry cost is constant. However, we do not see significantly higher interest rates with development across countries. For example, Caselli and Feyrer (2007) report an average return to capital of 8.4% for rich countries and 6.9% for poor countries. Based on levels of financial development, the more common assumption is that firms are more financially constrained in developing countries, not developed ones.

On variations in post entry growth rates, Hsieh and Klenow (2014) do find that incumbents grow faster in the U.S. than in India and Mexico. But the U.S. seems to be an outlier in this regard, judging from evidence on other countries in Scarpetta et al. (2002) and for five more countries documented by Hsieh and Klenow (2014). Finally, we do not see sufficient variation in the exit rate in Scarpetta et al. (2002) or in Hsieh and Klenow (2014).[16] Therefore, it is unlikely that our cross-country empirical findings are driven by variations in the discount rate, growth rate and exit rate.

Similarly, we do not observe the magnitude and direction of variation in the discount rate, post-entry growth rate and exit rate across time within a country needed to explain the variation in average revenue over time. U.S. manufacturing establishments' exit rate declined by about 2 percentage points from 12% between 1975 and 2010. The U.S. establishment post-entry growth rate appears stable in Hsieh and Klenow (2014). If the entry cost was constant over time in the U.S., therefore, then to explain the rise in average revenue the discount rate must be increasing. However, if anything real interest rates have fallen in the U.S. from the 1980s through the latest data.

---

[16]For example, if entry costs are the same in India and the U.S., $\sigma = 4$, exit rates in the U.S. and India are 0.1 and 0.05, post-entry growth rates in the U.S. and India are 0.02 and 0.003, and the U.S. interest rate is 0.05, then India's interest rate would have to be negative to explain its small average revenue relative to the U.S. Here δ and g are picked to fit the empirical patterns in Hsieh and Klenow (2014).

5.3. Measurement error in output

The modest relationship we find between plants per worker and value added per worker across time and space could be biased downward by measurement error in value added. In Section 3 we showed that the results extend to looking only at OECD countries, and over time in the U.S. Measurement error should be less of a concern in these samples. Moreover, measurement error would have to swamp the signal to explain our results. Consider the slope of 0.20 in Figure 5. Even if one-half of the variation in Y across countries stems from classical measurement error, the true elasticity of plants per worker with respect to value added per worker would be only 0.40. And this ignores any classical measurement error in employment, which would bias the elasticity upward.

5.4. Labor share in goods production

Production uses capital, not just labor, so value added per worker may not be a good proxy for the real wage and the cost of labor used for entry. If rich countries have higher labor shares, we could be overestimating the elasticity of entry costs with the real wage because the real wage does not vary in full proportion to value added. We entertained this possibility by controlling for labor share using the values from Caselli and Feyrer (2007). We find that our slope estimates are robust to this.

6. Conclusion

In manufacturing, the number of plants or firms per worker increases modestly with output per worker. This is true over time in the U.S., China and India, and across countries inside and outside the OECD. It is true within industries, not just across them. The number of businesses is more closely tied to the number of workers.

These facts can be explained by a model in which entry costs rise with labor

productivity. Entry costs can rise with productivity for multiple reasons. First, if entry is labor-intensive then higher wages that go along with higher labor productivity raise the cost of entry. Second, the costs of setting up operations could be increasing with the level of technology, worker skill, or physical capital per worker. We leave it for future research to try to distinguish between these explanations.

We draw out several implications for policy and modeling. First, policies that boost productivity need not increase the number of firms or plants. Second, if the choice is between denominating entry costs in terms of labor or output, the more realistic choice is fixed entry costs in terms of labor. Third, we empirically corroborate the common assumption in endogenous growth models that the cost of innovation rises with the level of technology attained. Fourth, technological entry costs appear to be at least as important as the government-imposed costs captured by the World Bank's Doing Business Survey.

7.   **Appendix**

We compared our measure of entry costs with other alternative measures of regulatory entry costs in Barseghyan and DiCeccio (2011). We find that regulatory entry costs do not explain a large part of the rise in entry costs.
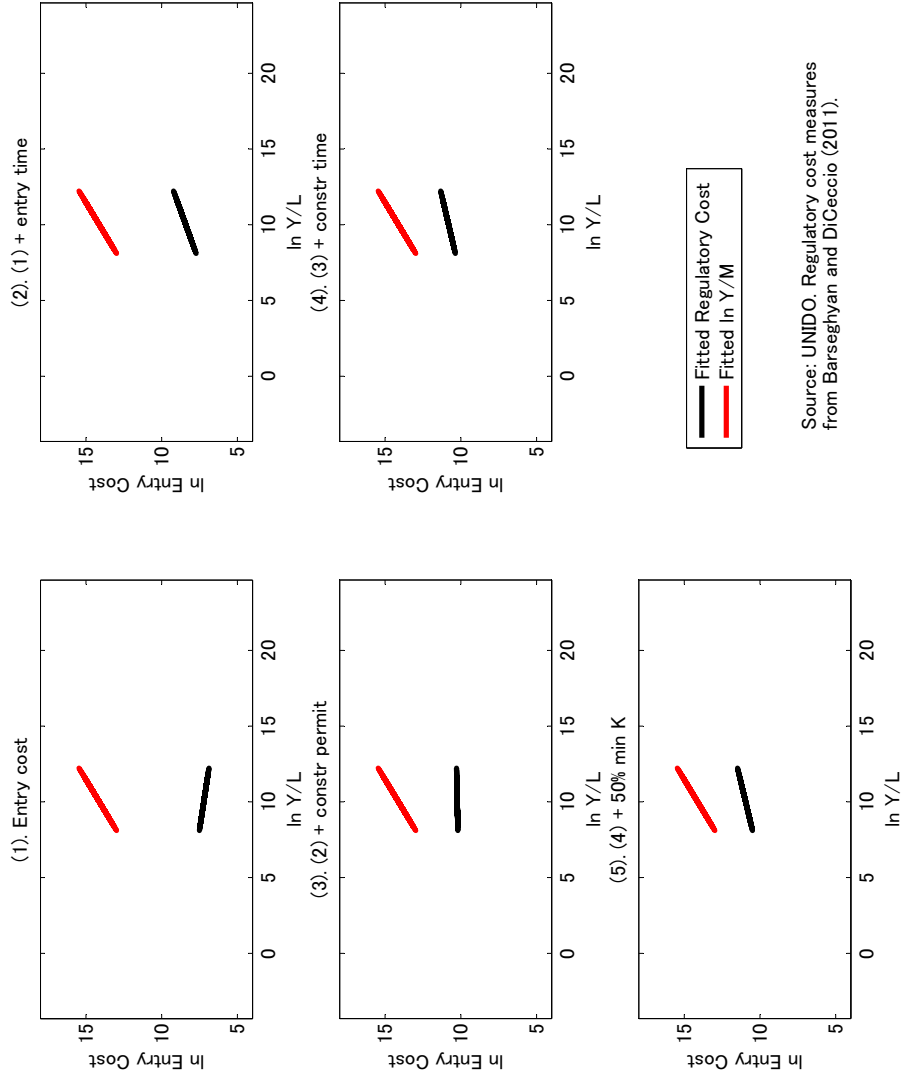
Figure 10: Overall entry costs vs. broader measures of regulatory entry costs.

References

Acemoglu, Daron, *Introduction to Modern Economic Growth*, Vol. 1, Princeton University Press, 2011.

—, Ufuk Akcigit, Nicholas Bloom, and William Kerr, "Innovation, Reallocation and Growth," April 2013.

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), pp. 323–351.

Atkeson, Andrew and Ariel Toms Burstein, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 2010, *118* (3), pp. 433–484.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multi-Product Firms and Product Switching," *The American Economic Review*, 2010, *100* (1), 70–97.

Bhattacharya, Dhritiman, Nezih Guner, and Gustavo Ventura, "Distortions, endogenous managerial skills and productivity differences," *Review of Economic Dynamics*, 2012.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

Bresnahan, Timothy F. and Peter C. Reiss, "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 1991, *99* (5), pp. 977–1009.

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

— and —, "Product Creation and Destruction: Evidence and Implications," *The American Economic Review*, 2010, *100* (3), 691–723.

Caselli, Francesco and James Feyrer, "The Marginal Product of Capital," *The Quarterly Journal of Economics*, 2007, *122* (2), pp. 535–568.

Costinot, Arnaud and Andrés Rodríguez-Clare, *Trade Theory with Numbers: Quantifying the Consequences of Globalization*, Vol. 4, 2013.

David, Joel M, "The Aggregate Implications of Mergers and Acquisitions," 2013.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The Regulation of Entry," *The Quarterly Journal of Economics*, 2002, *117* (1), pp. 1–37.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, markups, and the gains from international trade," Technical Report, National Bureau of Economic Research 2012.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, September 2008, *98* (1), 394–425.

Grossman, Gene M and Elhanan Helpman, "Quality ladders in the theory of growth," *The Review of Economic Studies*, 1991, *58* (1), 43–61.

Hopenhayn, Hugo A, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica*, 1992, pp. 1127–1150.

Hsieh, Chang-Tai and Peter J. Klenow, "Relative Prices and Relative Prosperity," *The American Economic Review*, 2007, *97* (3), 562–585.

＿ and ＿ , "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 2014. Forthcoming.

Klette, Tor Jakob and Samuel Kortum, "Innovating firms and aggregate innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, pp. 1389–1419.

Loecker, Jan De, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik, "Prices, markups and trade reform," Technical Report, National Bureau of Economic Research 2012.

Lucas, Robert E Jr., "On the size distribution of business firms," *The Bell Journal of Economics*, 1978, pp. 508–523.

Luttmer, Erzo GJ, "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 2007, *122* (3), 1103–1144.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), pp. 1695–1725.

Melitz, Marc J and Gianmarco IP Ottaviano, "Market size, trade, and productivity," *The Review of Economic Studies*, 2008, *75* (1), 295–316.

Porta, Rafael La and Andrei Shleifer, "The Unofficial Economy and Economic Development," *The Brookings Papers on Economic Activity*, 2008, pp. 275–352.

Rivera-Batiz, Luis A and Paul M Romer, "Economic integration and endogenous growth," *The Quarterly Journal of Economics*, 1991, *106* (2), 531–555.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), pp. S71–S102.

Romer, Paul M, "The origins of endogenous growth," *The Journal of Economic Perspectives*, 1994, *8* (1), 3–22.

Scarpetta, Stefano, Philip Hemmings, Thierry Tressel, and Jaejoon Woo, "The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data," April 2002.

Tybout, James R., "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?," *Journal of Economic Literature*, 2000, *38* (1), pp. 11–44.