Entry Costs Rise with Growth

Peter J. Klenow* Huiyu Li†

March 11, 2024

Click here for the latest version

Abstract

Over time and across states in the U.S., the number of firms is more closely tied to overall employment than to output per worker. In many models of firm dynamics, trade, and growth with a free entry condition, these facts imply that the costs of creating a new firm increase sharply with productivity growth. This increase in entry costs can stem from the rising cost of labor used in entry and weak or negative knowledge spillovers from prior entry. How entry costs vary with growth matters for welfare. For example, our findings suggest that productivity-enhancing policies will not induce entry of firms, thereby limiting the total impact of such policies on welfare.

JEL Classification: E23, O47

---

*klenow@stanford.edu Department of Economics, Stanford, SIEPR, and NBER.

†huiyu.li@sf.frb.org Economic Research, Federal Reserve Bank of San Francisco. This project arose from a previous version with Albert Bollard entitled "Entry Costs Rise with Development." We thank research assistants Nathaniel Barlow, Patrick Kiernan, Jack Mueller, and Sam Tarasewicz for their excellent work. We are grateful for helpful comments from our discussant Ariel Burstein and from seminar participants at Stanford, the Federal Reserve Bank of San Francisco, the University of Washington, the Federal Reserve Bank of Philadelphia, SED Madison, the NBER Summer Institute, and the Federal Reserve Bank of St. Louis. Opinions and conclusions herein are those of the authors and do not necessarily represent the views of the Federal Reserve System, the Federal Reserve Bank of San Francisco, or the U.S. Census Bureau. This research was performed at a Federal Statistical Research Data Center under FSRDC Project 1440. All results have been reviewed by the U.S. Census Bureau to ensure that no confidential information is disclosed. The authors have no relevant or material financial interests that relate to the research in this paper.

1 Introduction

Suppose that new businesses are created with a fixed amount of output. Then a policy which boosts productivity can generate an endogenous expansion in the number of firms, with gains in welfare to the extent more firms entails more varieties. This multiplier effect through entry is analogous to the multiplier effect on output from physical capital accumulation in the neoclassical growth model. If instead entry requires a fixed amount of labor, however, then policies boosting productivity will fail to generate additional 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. Some models assume entry costs are stable or stationary (e.g. a fixed output cost to invent a new product).[1] Other models assume entry costs rise as growth proceeds, say because entry requires a fixed amount of labor and labor becomes more expensive with growth.[2] Some studies do not take a stand but emphasize that the entry technology matters for the welfare impact of policies.[3]

Entry costs may also depend on knowledge spillovers from past entry. In the growth literature, it is common to assume spillovers from previous innovation to future innovation. This includes the classic models of Romer (1990) and Aghion and Howitt (1992) as well as many successors. Jones (1995) and Bloom, Jones, Van Reenen and Webb (2020) argue that such spillovers are limited or even negative.

[1] See Hopenhayn (1992), Hopenhayn and Rogerson (1993), Foster, Haltiwanger and Syverson (2008), Clementi and Palazzo (2016), Gutierrez, Jones and Philippon (2019), David (2021), Boar and Midrigan (2022a,b), and Karahan, Pugsley and Şahin (2024).

[2] Examples include Lucas (1978), Grossman and Helpman (1991), Melitz (2003), Klette and Kortum (2004), Luttmer (2007), Bilbiie, Ghironi and Melitz (2012), Acemoglu, Akcigit, Alp, Bloom and Kerr (2018), Atkeson and Burstein (2019), Sterk, Sedlek and Pugsley (2021), Hopenhayn, Neira and Singhania (2022), Baqaee, Farhi and Sangani (2023), and Peters and Walsh (2022).

[3] For example, Rivera-Batiz and Romer (1991), Atkeson and Burstein (2010), Bhattacharya, Guner and Ventura (2013), survey by Costinot and Rodríguez-Clare (2014), and Baqaee and Farhi (2021).

Existing evidence is limited on how entry costs change with growth. The overall distribution of employment across firms and plants provides some indirect evidence. Laincz and Peretto (2006) report no trend in U.S. average firm employment. Luttmer (2007, 2010) shows that entry costs proportional to average productivity are consistent with a stationary firm size distribution in various growth models. While our paper studies the secular trend in entry costs, Karahan, Pugsley and Şahin (2023) use the free entry condition to infer whether entry costs are cyclical, and Gutiérrez and Philippon (2019) focus on the cross-industry relationship between the entry rate and Tobin's Q. Bento and Restuccia (2022) incorporate data on nonemployer establishments, and show this affects inference about trends in average employment per firm.

In this paper, we provide evidence on how the average employment per firm varies with the level of overall labor productivity. We look over time and across states in the Business Dynamics Statistics (BDS) maintained by the U.S. Census, in particular from 1978 through 2020. We combine this Census data with U.S. Bureau of Economic Analysis (BEA) data on aggregate and state labor productivity. We argue that these simple empirical elasticities discipline the nature of entry costs in widely used models.

We find that average employment per firm is stable or increases with the level of labor productivity, both over time and across states. These patterns imply that revenue per firm increases sharply with growth. Firms evidently need more revenue to satisfy the free entry condition in places and times with higher market-wide labor productivity. If higher revenue is associated with higher operating profits, then entry costs must be bigger in order for the zero profit condition to hold. We consider other possible explanations, however, such as trends in firm markups, exit rates, post-entry growth rates, discount rates, selection, and industry composition. We will argue that these competing forces are too weak to explain why average employment per firm does not decline significantly relative to the extent of labor productivity growth.

We illustrate the implications of our empirical findings using two deliberately simple and stylized models. One model features long run growth at the country level. The second model contains growing U.S. states with mobility of workers and firms. In these models, entry costs can rise with growth simply because entry is labor-intensive and labor becomes more expensive when productivity grows. Entry costs could also rise with growth because it is more costly for entrants to set up more technologically sophisticated operations as the economy advances (say due to limited or negative knowledge spillovers).[4] We use our empirical findings to estimate parameters governing the labor-intensity of entry costs and the relationship between entry costs and the level of technology. We find that fitting our facts requires that entry be labor-intensive and/or that knowledge spillovers are weak, thereby explaining why entry costs rise with growth.

We draw the following three conclusions 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 at best weak knowledge spillovers for innovation embodied in entry. Third, productivity-enhancing policies have muted effects on entry, and hence are not amplified through endogenous entry.

The rest of the paper proceeds as follows. Section 2 provides two models to illustrate why we care about the nature of entry costs and to motivate our empirical design. Section 3 presents evidence on how the number of businesses varies with growth over time and across states in the U.S. and draws potential implications for entry costs. Section 4 estimates entry technology parameters and discusses the welfare implications. Section 5 gauges the robustness of our empirical findings and Section 6 concludes.

[4]Our evidence is relevant for total entry costs, which are the sum of technological and regulatory entry costs. In the *Doing Business* surveys, regulatory costs of entry (relative to GDP per capita) fall with development, as shown by Djankov, Porta, Lopez-de Silanes and Shleifer (2002). Thus rising technological entry costs with development may be needed to explain why employment per firm is higher in richer countries, as documented by Bento and Restuccia (2017).

2 Simple motivating models

We first present a stylized love-of-variety model of a one-region economy to illustrate how the elasticity of entry costs with respect to growth matters for welfare. Then we extend the model to multiple regions à la Redding and Rossi-Hansberg (2017) and Redding (2022) to guide our cross-state empirical analysis. As both models are standard, we relegate the details on them to Appendix A. These models assume that the number of varieties is proportional to the number of firms. This is assumed in many other models, such as Peters and Walsh (2022). In Appendix B we lay out a model with endogenous varieties per firm in which we recover the same estimating equation as we derive below.[5]

2.1 One-region model

We first 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 proportional 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 with elasticity of substitution σ. 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.[6] 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

[5]We show in Appendix B that entry costs matter for welfare in other models, such as a version of the static Lucas span-of-control model, a static love-of-variety model with congestion in contemporaneous entry, and an endogenous growth model with expanding varieties per firm (rather than a single product per firm as in our baseline model).

[6]We could allow post-entry heterogeneity in firm technology and define $A := (\mathbb{E}A_f^{\sigma-1})^{\frac{1}{\sigma-1}}$, where A_f is firm-level productivity.

between the wage bill, revenue, and profit in each firm

$$wl = \frac{\sigma - 1}{\sigma} \cdot py = (\sigma - 1) \cdot \pi \tag{1}$$

Let L^y be the total amount of labor devoted to producing intermediate goods and N be the total number of intermediate goods produced. By symmetry of the intermediate goods production function, aggregate output is given by

$$Y = A \cdot L^y \cdot N^{\frac{1}{\sigma-1}} \tag{2}$$

One unit of an entry good is required to create a variety, which is the equivalent to setting up an intermediate 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, 2015) in assuming that the entry technology has the Cobb-Douglas form

$$N = A^e \cdot (Y^e)^{1-\lambda} \cdot (L^e)^{\lambda} \tag{3}$$

where L^e and Y^e are the amount of labor and final output, respectively, used in creating varieties.

This specification of the entry technology nests various assumptions in the literature. For example, entry costs are as in (3) but with $\lambda = 1$ and $A^e_s = 1$ in Lucas (1978), Romer (1990), Melitz (2003), Luttmer (2007), and Hopenhayn et al. (2022). When $\lambda = 0$ and $A^e_s = 1$, entry costs are as in Hopenhayn and Rogerson (1993), Foster, Haltiwanger and Syverson (2008), David (2021), and Karahan, Pugsley and Şahin (2024). Finally, entry costs may rise with labor productivity if, as in Berry and Waldfogel (2010), Cole, Greenwood and Sanchez (2016), and Bento and Restuccia (2017), better production technologies carry higher setup costs (lower A^e).

Perfect competition in the CRTS sector producing entry goods implies that the equilibrium cost and price of creating a variety in terms of consumption goods satisfy

$$p^e \propto \frac{w^\lambda}{A^e} \tag{4}$$

and the labor share of revenue in entry goods production is

$$\frac{wL^e}{p^e N} = \lambda \tag{5}$$

Free entry into producing intermediate goods firms (and varieties), with positive entry in equilibrium, implies profit per variety equals to the entry cost

$$\pi = p^e \tag{6}$$

Thus the one-shot equilibrium, given the triple $\{L, A, A^e\}$, consists of prices $\{w, p^e\}$ and allocations $\{C, N, Y, L^e, L^y\}$ such that (1) to (6) hold, and the following labor and goods market clearing conditions are satisfied:

$$L = L^y + L^e, \quad Y = C + Y^e$$

We now consider how the welfare impact of a change in intermediate goods productivity A depends on the entry technology. In equilibrium, welfare (equivalently, the real wage) increases A and the number of varieties

$$w = \frac{\sigma - 1}{\sigma} \cdot A \cdot N^{\frac{1}{\sigma-1}}$$

and the change in welfare from a change in A is

$$\frac{\partial \ln w}{\partial \ln A} = 1 + \frac{1}{\sigma - 1} \cdot \frac{\partial \ln N}{\partial \ln A}$$

An increase in A not only raises welfare directly (the first term, or 1), but also has the potential to improve welfare indirectly through variety expansion (the second term).

One can show that equilibrium variety satisfies

$$N \propto \frac{wL}{p^e}$$

such that the number of varieties depends on the value of labor relative to the entry cost. Combining this with equation (4) relating the real wage to p^e, we get

$$\frac{\partial \ln N}{\partial \ln A} = (1-\lambda) \cdot \frac{\partial \ln w}{\partial \ln A}$$

That is, the elasticity of variety with respect to A is larger when the share of output used in producing varieties $(1-\lambda)$ is bigger. Higher A means more output, and some of this output is devoted to producing more varieties if the final good is used in entry $(\lambda < 1)$. Incorporating this channel, the total impact of A on welfare is

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

with the second term capturing the effect of variety expansion. A higher output share $(1-\lambda)$ means more amplification.

The amplification of an increase in productivity depends negatively on the elasticity of substitution σ, 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 product level. For $\sigma = 4$, the increase in the total impact relative to the direct impact ranges from 50% when $\lambda = 0$ to 0% when $\lambda = 1$. Thus, for a plausible value of σ, the nature of entry costs matters immensely for the welfare impact of changes in A from technology or allocative efficiency.

The entry technology also influences the welfare impact of policies that affect the level of the population. As in Melitz (2003), increasing the population is like going from autarky to frictionless trade between two symmetric countries. In this case, the overall welfare effect is

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

Just as with an increase in A, for an increase in L the amplification through variety expansion is 50% when $\lambda = 0$ and 0% when $\lambda = 1$.

2.2 Spatial model

Next we extend the simple one region model to multiple regions. This allows us to speak to evidence on changes in firm size not just at the national level, but also at the state level. We also generalize the entry technology to allow knowledge spillovers from past entry and endogenize the growth rate of A. We view the cross-state evidence as more credible given that we can control for national trends in markups, firm age composition, etc.

2.2.1 Environment

The economy consists of $s = 1, 2, \ldots, S$ states and an exogenous mass of identical workers L. Each worker chooses one state to live in and to supply one unit of labor to the firms in that state. Ex-ante identical firms choose one state in which to produce. The mass of workers living in each state L_s and the mass of firms in each state N_s are therefore endogenous. States differ in their endowment of housing H_s, intermediate goods productivity A_s, and entry efficiency A_s^e. Intermediate goods sent from state s' to state s incur an iceberg trade cost denoted by $d_{s,s'} > 1$ if $s \neq s'$ and $d_{ss} = 1$. We assume the trade cost is symmetric ($d_{s,s'} = d_{s',s}$).

The government owns the housing in each state. They set rent r_s for each unit of housing so that all available housing is used. Rents are then redistributed to each worker residing in the state as lump sum payment τ_s. The workers in state s own the firms in state s and receive equal shares of firm profits net of entry costs $(\pi_s - p_s^e)N_s/L_s$.

2.2.2 Final goods production

In each state s, final goods are produced using the CES technology

$$Y_s = \left[\sum_{s'=1}^{S} \int_0^{N_{s'}} y_{s,s'}(j)^{\frac{\sigma-1}{\sigma}}\, dj\right]^{\frac{\sigma}{\sigma-1}}$$

where $y_{s,s'}(j)$ is the quantity of intermediate input variety j produced by firm j in state s' and sold to state s.

Let $p_{s,s'}(j)$ denote the price of this good in state s. Profit maximization by perfectly competitive final goods producers implies that the price of the final good in state s is

$$P_s = \left[\sum_{s'=1}^{S} \int_0^{N_{s'}} p_{s,s'}(j)^{1-\sigma}\, dj\right]^{\frac{1}{1-\sigma}}$$

and demand for each variety in state s is given by

$$\frac{y_{s,s'}(j)}{Y_s} = \left(\frac{p_{s,s'}(j)}{P_s}\right)^{-\sigma}$$

2.2.3 Worker's problem

The utility of a worker in state s is a Cobb-Douglas combination of consumption of the final good and housing:

$$U_s = \left(\frac{c_s}{\alpha}\right)^{\alpha} \cdot \left(\frac{h_s}{1-\alpha}\right)^{1-\alpha}, \quad \alpha \in (0,1)$$

The worker maximizes U_s by choosing c_s and h_s subject to the budget constraint

$$P_s c_s + r_s h_s \le w_s + (\pi_s - p_s^e)\, N_s/L_s + \tau_s =: v_s$$

The consumer spends α share of their income v_s on consumption and the rest on housing:

$$P_s c_s = \alpha \cdot v_s, \quad r_s h_s = (1-\alpha) \cdot v_s$$

Workers choose to live and work in the state that gives them the highest utility.

2.2.4 Entry technology

To produce in state s, a firm buys an entry good that is produced using local labor l_s^e and the state's final consumption good y_s^e according to the Cobb-Douglas technology

$$N_s = e^{\epsilon_s} \cdot N_{s,-1}^{\phi} \cdot \left(\frac{l_s^e}{\lambda}\right)^{\lambda} \left(\frac{y_s^e}{1-\lambda}\right)^{1-\lambda}, \quad \lambda \in (0,1).$$

In terms of the previous model's notation, overall entry efficiency is $A_s^e = e^{\epsilon_s} \cdot N_{s,-1}^{\phi}$, where $N_{s,-1}^{\phi}$ captures spillover from the past stock of varieties and ϵ_s captures other factors affecting the efficiency of entry goods production. When $\phi > 0$, entry efficiency increases with the stock of varieties.

As before, we assume the market for entry goods is perfectly competitive, so the equilibrium price of the entry good p_s^e/P_s increases with factor prices and declines with entry efficiency A_s^e:

$$\frac{p_s^e}{P_s} \propto \left(\frac{w_s}{P_s}\right)^{\lambda} \frac{1}{A_s^e} \tag{7}$$

2.2.5 Intermediate goods firm's problem

Intermediate goods producers in state s are ex-ante identical and have the same productivity A_s after entry into state s. As a result, producers in each state make the same decision and we drop the firm j index. A firm in state s can produce y units of its variety using y/A_s units of labor. Since delivering a unit of the good from state s' to state s requires $d_{s,s'}$ units of the good, the labor input needed by a firm in state s' to deliver y units of goods to state s is given by

$$l_{s,s'} = y \cdot \frac{d_{s,s'}}{A_{s'}}$$

Given this technology and the demand function in each state s, a firm in state s' chooses prices $p_{s,s'}$ for each destination state s to maximize post-entry profits

$$\sum_{s=1}^{S} \left(p_{s,s'} - w_{s'} \frac{d_{s,s'}}{A_{s'}} \right) \left(\frac{p_{s,s'}}{P_s} \right)^{-\sigma} Y_s$$

The optimal price is a fixed markup over the marginal cost, where the firm charges a higher price for destinations with larger trade costs:

$$p_{s,s'} = \frac{\sigma}{\sigma - 1} \cdot \frac{d_{s,s'}\, w_{s'}}{A_{s'}}$$

The profit for selling to state s is thus

$$\pi_{s,s'} = \frac{p_{s,s'}\, y_{s,s'}}{\sigma}$$

and a firm enters in state s' if and only if its total profits across all destinations covers the entry cost:

$$\pi_{s'} := \sum_{s=1}^{S} \pi_{s,s'} \geq p^e_{s'}$$

2.2.6 Closing the model

Given L and $\{A_s, A^e_s, H_s, d_{s,s'}\}$, an equilibrium consists of prices $\{w_s,\, r_s,\, P_s,\, p^e_s\}$ in each location s and $p_{s,s'}$ for each trading pair (s, s'), and allocations $\{c_s, h_s, L_s, L^e_s, L^y_s, C_s, Y_s, Y^e_s, N_s, \tau_s, y_{s,s'}, l_{s,s'}\}$ such that for each state s

1. $\{c_s, h_s\}$ solve the worker's problem given prices and transfers

2. $\{l_{s,s'}, y_{s,s'}, p_{s,s'}\}$ solve the intermediate good firm's problem

3. $\{L^e_s, Y^e_s\}$ solve the entry goods producers problem

4. the zero profit condition for intermediate good producers holds:
   $N_s(\pi_s - p_s^e) = 0, \quad \pi_s - p_s^e \geq 0, \quad N_s \geq 0$

5. land markets clear: $H_s = L_s h_s$

6. labor markets clear: $L_s = L_s^e + L_s^y$ and $L = \sum_s L_s$

7. final goods markets clear: $Y_s = C_s + Y_s^e$, where $C_s = L_s c_s$

8. government budgets are balanced: $r_s H_s = \tau_s L_s$

9. workers are indifferent between locations

Since the model is standard, we refer readers to Appendix A for the solution of the model. Next, we turn to how entry technology parameters modulate the welfare effects of changes in productivity.

2.2.7 Entry and shocks to the level of productivity

Welfare depends on consumption and housing. Consumption is equal to the real wage in the equilibrium $c_s = w_s/P_s$ because of the household's budget constraint, the zero profit condition for intermediate goods firms and the balanced government budget condition. The real wage in turn is equal to

$$\ln \frac{w_s}{P_s} = \text{constant} + \frac{\ln A_s^e + \ln L + (\sigma - 1)\ln A_s + \ln \frac{L_s}{L} + (\sigma - 1)\ln(n_{s,s}) - \sigma \ln(b_{s,s})}{\sigma - 1 - (1 - \lambda)} \tag{8}$$

where $b_{s,s}$ is the expenditure share in state s on local goods and $n_{s,s}$ is the share of production labor used to produced domestically-consumed goods.

For illustration, consider symmetric states with the same initial values of $\{A_s, A_s^e, H_s, d_{s,s'}\}$. And consider a common change in A_s to clarify the model's properties. In this case, L_s/L, $n_{s,s}$ and $b_{s,s}$ do not change. As in the one-region model, the elasticity of the real wage in every state to the A_s shock is

$$\frac{\partial \ln w_s/P_s}{\partial \ln A_s} = 1 + \frac{1 - \lambda}{\sigma - 1 - (1 - \lambda)}$$

while the elasticity with respect to L and a common A_s^e shock is

$$\frac{\partial \ln w_s/P_s}{\partial \ln A_s^e} = \frac{\partial \ln w_s/P_s}{\partial \ln L} = \frac{1}{\sigma-1}\left(1+\frac{1-\lambda}{\sigma-1-(1-\lambda)}\right).$$

Since H_s/L_s does not change in response to shocks to the common values for A_s and A_s^e, the elasticity of consumption-equivalent welfare is the same as the elasticity of the real wage with respect to an A_s or an A_s^e shock. Consumption per capita increases when total population increases. However, housing per capita also declines. Hence the consumption-equivalent welfare impact of a shock to total population L is

$$\frac{1}{\sigma-1-(1-\lambda)} - \frac{1-\alpha}{\alpha}$$

To recap, a smaller labor share in entry (lower λ) amplifies the positive effects on welfare from higher productivity, entry efficiency, or population.

2.2.8 Entry and amplification of productivity growth

The above model describes an one-shot economy where state productivity A_s is given. A simple way to introduce endogenous growth in A_s is to let each firm j in state s choose its productivity $A_{st}(j)$ in a way that builds upon aggregate productivity in the previous period $A_{s,t-1}$. As we show below, both the labor share in entry λ and the spillover from past varieties ϕ are important for the effect of increases in the growth rate of A_s on the *growth* rate of the economy.

In addition to past entry, let the entry efficiency A_{st}^e also depend on the productivity the firm chooses relative to past aggregate productivity and a shock to entry efficiency that is common to all firms in state s:

$$\ln A_{st}^e = \phi \ln N_{s,t-1} - \mu \cdot \frac{A_{s,t}(j)}{A_{s,t-1}} + \epsilon_{st}$$

A positive μ means that entry costs increase with the productivity chosen by the entering firm, $A_{s,t}(j)/A_{s,t-1}$. In each period, the firms observe the common entry efficiency shock ϵ_{st} and then decide $A_{st}(j)$. As before, entry costs in equilibrium are given by

$$\frac{p^e_{st}}{P_{st}} = \frac{\left(\frac{w_{st}}{P_{st}}\right)^\lambda}{A^e_{st}}$$

Profit maximization by intermediate goods producers and free entry imply that the choice of $A_{st}(j)$ by firm j satisfies[7]

$$\frac{\partial \ln \pi_{st}(j)}{\partial \ln A_{st}(j)} = \frac{\partial \ln p^e_{st}(j)}{\partial \ln A_{st}(j)}$$

Since variable profits $\pi_{st}(j)$ are proportional to $A_{st}(j)^{\sigma-1}$, the firm's optimal choice of $A_{st}(j)$ is given by

$$\sigma - 1 = \mu \cdot \frac{A_{st}(j)}{A_{s,t-1}}$$

and all regions have the same growth in A:

$$g^A_t := \ln \frac{A_{st}(j)}{A_{s,t-1}} = \ln \frac{\sigma - 1}{\mu}$$

The growth rate increases with the elasticity of substitution σ and declines with the elasticity μ of entry costs with respect to growth in A. Following from this, entry efficiency in equilibrium is $\ln A^e_{st} = \phi \ln N_{s,t-1} - (\sigma - 1) + \epsilon_{st}$.

At the equilibrium, the number of firms in each state grows at rate

$$g^N_{st} = \frac{g^L_t + (1-\lambda)\, g^{w/p}_t}{1-\phi} \tag{9}$$

[7]Firms choose $A_{s,t}(j)$ to maximize profit post-entry costs. Hence $A_{s,t}(j)$ satisfies the first order condition

$$\frac{\partial \pi_{st}(A_{st}(j))}{\partial A_{st}(j)} = \frac{\partial p^e_{st}(A_{st}(j))}{\partial A_{st}(j)}$$

At the equilibrium, we also have $\pi_{st}(A_{st}(j)) = p^e_{st}(A_{st}(j))$ and hence

$$\frac{\partial \ln \pi_{st}(A_{st}(j))}{\partial \ln A_{st}(j)} = \frac{\partial \pi_{st}(A_{st}(j))}{\partial A_{st}(j)} \frac{A_{st}(j)}{\pi_{st}(A_{st}(j))} = \frac{\partial p^e_{st}(A_{st}(j))}{\partial A_{st}(j)} \frac{A_{st}(j)}{p^e_{st}(A_{st}(j))} = \frac{\partial \ln p^e_{st}(A_{st}(j))}{\partial \ln A_{st}(j)}.$$

which in turn implies that the real wage grows at rate[8]

$$g_{st}^{w/p} = \frac{(\sigma - 1)\, g_t^A + \frac{g_t^L}{1-\phi}}{\sigma - 1 - \frac{1-\lambda}{1-\phi}} \tag{10}$$

The trend growth of the real wage in each state is driven by the endogenous, common growth of state productivity g^A and by growth in the national population g^L. The wage effects of these driving forces are amplified through entry when λ is less than 1. The wage effects can be amplified when there are positive spillovers from the past variety stock to the efficiency of creating new varieties ($\phi > 0$), or dampened when there are negative spillovers ($\phi < 0$). The intuition is similar to the multiplier effect we detailed previously for the λ channel.

3 Evidence on entry costs and growth

Motivated by the previous section, we next consider what values of λ and ϕ are consistent with data on the number and productivity of U.S. firms. The free-entry condition is a zero-profit condition which equalizes average firm profits with the entry cost. Hence we can look at the relationship between firm profits and labor productivity over time and across states to infer how entry costs correlate with labor productivity. Another way is to look at the relationship between average employment per firm and labor productivity if the ratio of average payroll to average profit does not vary systematically with labor productivity. In this section, we examine the relationship between average employment and labor productivity because firm employment data is available for all industries. In Section 5, we show that our findings are robust to trends in the ratio of average payroll to average profit. We also directly examine

[8]Population and domestic expenditure shares by state also affect the real wage. These variables reflect heterogeneous entry efficiency ϵ_s, trade costs $d_{s,s'}$, and amenities H_s. Equation (10) holds these variables fixed over time. See Appendix A for details of the derivation.

the relationship between firm profits and labor productivity in manufacturing using restricted data from the Census of Manufacturing.

3.1 Strategy for estimating how entry costs vary with growth

From the free entry condition and the solution to the firm's problem, we can derive the following equilibrium relationship between the average payroll of intermediate goods producer and the entry cost in each state:

$$\frac{w_s}{P_s} \cdot \frac{L_s^y}{N_s} = (\sigma - 1)\frac{\pi_s}{P_s} = (\sigma - 1)\frac{p_s^e}{P_s} = (\sigma - 1)\left(\frac{w_s}{P_s}\right)^{\lambda} \frac{1}{A_s^e} \tag{11}$$

The first equality comes from production worker payroll per firm being proportional to firm profits, while the second equality comes from the free entry condition. The last equality derives from the entry technology that links entry costs with the real wage and entry efficiency.

Rearranging (11), we can look at how employment per firm varies with the real wage to infer how entry costs vary with the real wage:

$$\frac{L_s^y}{N_s} = (\sigma - 1)\left(\frac{w_s}{P_s}\right)^{\lambda - 1} \frac{1}{A_s^e} \tag{12}$$

If entry uses only labor ($\lambda = 1$) and entry efficiency is constant, entry costs increases one for one with the real wage. Through the free entry condition, this implies that firm profits and hence firm payroll likewise increase one for one with the real wage. Since payroll is employment multiplied by the real wage, this further implies that employment per firm is invariant to changes in the real wage. In contrast, if entry uses only goods ($\lambda = 0$), then entry costs and payroll per firm are constant, which implies that employment per firm declines proportionately with the real wage.

As we will discuss later in this section, we have data on all workers (production and entry labor combined) and on gross state product (GSP). The model predicts a similar relationship between these data variables as it does for production workers and real wages in (12). If σ and λ are the same across states, then production workers per firm are proportional to total employment per firm:

$$L_s = L_s^e + L_s^y = \frac{\lambda N_s p_s^e}{w_s} + L_s^y = \left(\frac{\lambda}{\sigma - 1} + 1\right) L_s^y$$

We measure the real wage using local labor productivity GSP/L_s since

$$\frac{GSP_s}{L_s} = \frac{N_s}{L_s} \sum_{s'}^{S} \frac{p_{s',s} y_{s',s}}{P_s} = \frac{N_s}{L_s} \frac{w_s}{P_s} \sum_{s'}^{S} \frac{\sigma}{\sigma - 1} l_{s',s} = \frac{\sigma}{\sigma - 1} \frac{w_s}{P_s} \frac{L_s^y}{L_s} = \frac{w_s}{P_s} \frac{\sigma}{\sigma - 1 + \lambda}$$

Substituting the expressions for L_s and GSP_s/L_s into (12) yields the following equations involving observed variable and parameters:

$$\ln \frac{L_s}{N_s} = \text{constant} + (\lambda - 1) \ln \frac{GSP_s}{L_s} - \phi \ln N_{s,t-1} - \epsilon_s \tag{13}$$

Consistent with our growth model, we can also look at the relationship between changes in employment per firm and changes in GSP per worker:

$$\Delta \ln \frac{L_{st}}{N_{st}} = (\lambda - 1) \Delta \ln \frac{GSP_{st}}{L_{st}} - \phi \Delta \ln N_{s,t-1} - \Delta \epsilon_{st} \tag{14}$$

This equation holds even if the elasticity of substitution σ and hence the ratio of payroll to revenue varies across states.

We will run OLS regressions corresponding to (13) and (14) to show that employment per firm is stable relative to variations in GSP per worker and the lagged number of firms. From the perspective of our model, these patterns imply that entry costs rise with labor productivity both across states and over time within states.

Although these regressions address the key question of whether entry costs rise with growth, the OLS regression coefficients do not correctly identify the parameters λ and ϕ which determine exactly why entry costs rise with growth. This is because the regressors (GSP_s and N_s) are endogenous to the residuals (entry efficiency ϵ_s) in this regression, according to our model. In Section 4, we will use GMM to estimate the values of λ and ϕ in a model-consistent fashion.

3.2 Empirical patterns

We use Business Dynamics Statistics (BDS) from the U.S. Census Bureau on employment L_{st} and the number of firms or establishments N_{st} in each state. The models we described above feature a one-shot equilibrium which does not distinguish between new firms and incumbent firms. In the data, however, we apply our inference strategy to new firms separately from all firms. We use real gross value added from the Bureau of Economic Analysis (BEA) to calculate GDP_t and GSP_{st}. We describe the data in more detail in Appendix C.

3.2.1 National time-series evidence

Figure 1 displays the result of regressing log number of firms on log employment and log employment per firm on log GDP per worker, respectively, over time. The data is yearly from 1978 to 2020. Both bilateral relationships are strongly positive both economically and statistically. If we regress log firms on both log employment and log GDP per worker at the same time, the coefficient on log employment increases, whereas the coefficient on log GDP per worker becomes small and insignificant. If we add a linear time trend to this multivariate regression, the coefficient on employment increases further and that on GDP per worker becomes modestly positive and significant. These results are consistent with entry costs being more labor intensive than goods intensive.

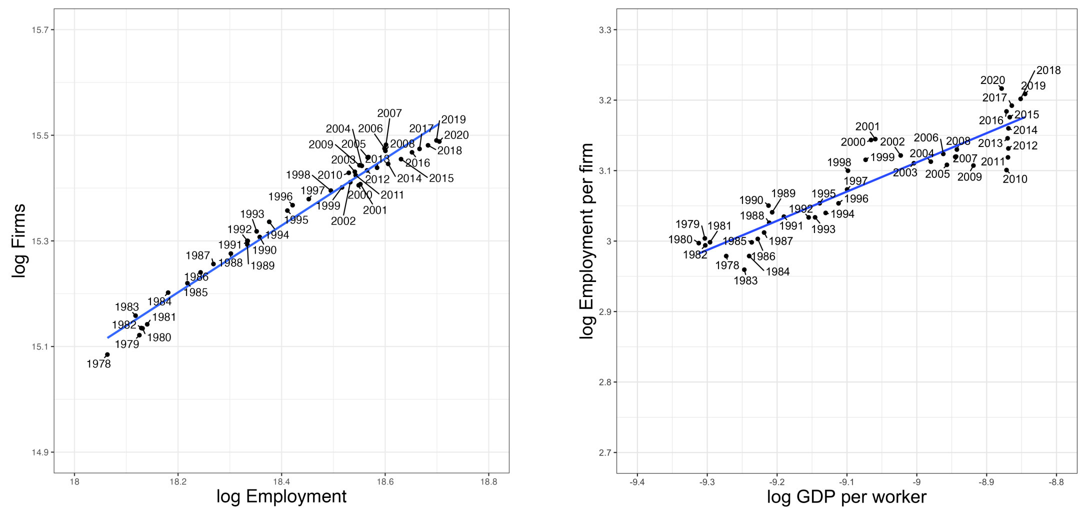
We now run OLS regressions based on (13) that is designed to explicitly get at how entry costs vary with growth. We add lagged firms to the regression in the spirit of the intertemporal knowledge spillovers (parameterized by ϕ) in our multi-region model.[9] Table 1 displays the result of regressing log employment per firm in the U.S. on log real GDP per worker and the lag of the log number

[9] The regression starts in 1979 when we add lagged number of firms because of the lag.

Figure 1: U.S. national time series, 1978–2020

(a) Number of firms vs. employment

(b) Employment per firm vs. GDP/employment



Notes: The number of firms and employment are from the U.S. Census Bureau's Business Dynamic Statistics (BDS). Real Gross Domestic Product (GDP) is from the U.S. Bureau of Economic Analysis (BEA). In the left panel, the slope coefficient is 0.63 (s.e. 0.02) and the R^2 = 0.98. For the right panel, the slope is 0.42 (0.03) and the R^2 is 0.85.

of firms at the national level.[10] Using non-overlapping 5-year averages generate similar results.

The first column of Table 1 displays the results when imposing $\phi = 0$, which is consistent with the one-region love-of-variety model. The second column also estimates ϕ. The regression using all firms and imposing $\phi = 0$ yields λ^{OLS} = 1.415 (s.e. 0.027) which implies an amplification of *negative* 12.1% (s.e. 0.7%). The second column shows λ^{OLS} = 1.236 (s.e. 0.073) and ϕ^{OLS} = -0.246 (s.e. 0.095), which implies that there was negative spillover from past entry. The amplification factor in this case remains mildly negative at -5.9% (s.e. 2.1%).

These regressions using data on all firms do not control for the aging of firms as documented by Karahan, Pugsley and Şahin (2024) and Hopenhayn, Neira and Singhania (2022). Since older firms tend to be larger, the average

[10]Table A2 in the Appendix displays the results using establishments instead of firms.

employment of firms may have risen due to aging rather than entry costs. The third and fourth columns of Table 1 run the same regression but using average employment of *new* firms as regressors, while keeping the explanatory variables the same.[11] We find that average employment of new firms and plants are stable relative to the rise in output per worker and rise in the number of firms. In column three where ϕ is restricted to be zero, we have λ^{OLS} = 0.907 (s.e. 0.055) and amplification = 3.2% (s.e. 1.9%). In column four, λ^{OLS} = 0.634 (s.e. 0.146), ϕ^{OLS} = -0.356 (s.e. 0.191) and amplification = 9.9% (s.e. 3.0%).

Table 1: Employment per firm on GDP per worker and lagged number of firms

National sample, 1978–2020

Dep variable	All firms	All firms	New firms	New firms
λ^{OLS}	1.415	1.236	0.907	0.634
	(0.027)	(0.073)	(0.055)	(0.146)
ϕ^{OLS}		-0.246		-0.356
		(0.095)		(0.191)
R^2	0.847	0.864	0.066	0.170
N	43	42	43	42
Amplification	-12.1%	-5.9%	3.2%	9.9%
	(0.7%)	(2.1%)	(1.9%)	(3.0%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

In sum, all the regressions using national data show that over the past four

[11] Using new firms, as opposed to all firms, controls for changes in the discount factor, post-entry growth rate and exit rate. We clarify this in section 5.1.

decades in the U.S., average employment per firm has been increasing and average employment per new firm has been stable while labor productivity grew. The free-entry condition in our baseline model interprets this pattern as a rise in entry costs with labor productivity so that amplification is modest. As mentioned, however, these OLS estimates are not model-consistent in that the residual reflect trends in entry costs that should affect the regressors. We carry out model-consistent GMM regressions in the next section.

3.2.2 State panel evidence

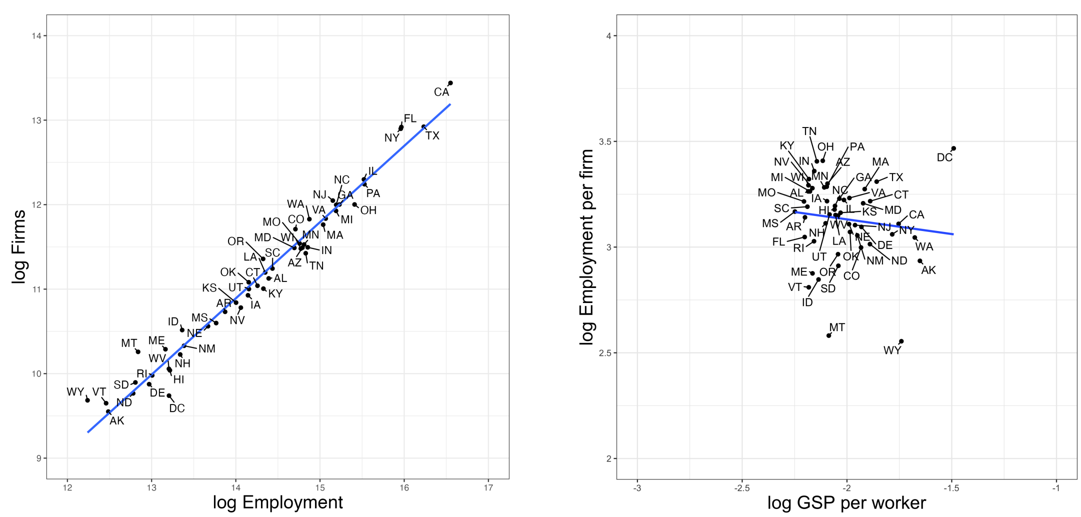
Figure 2 displays the result of regressing log number of firms on log employment and log employment per firm on log GSP per worker, respectively, across U.S. states in 2020, the latest year of the BDS data. The number of firms in a state is strongly and positively related to the number of workers in the state, but employment per firm is not related to GSP per worker in the state. If we regress log firms on both log employment and log GDP per worker, the coefficient on log employment is unaffected, whereas the coefficient on log GDP per worker remains small and insignificant. These patterns hold for other years as well. They are consistent with entry costs being denominated in terms of labor rather than goods.

Our spatial model has predictions for the cross-state relationship between changes in state level average firm size and the growth in real state output per worker — regression equation (14). Table 2 displays the OLS regression results when we regress the change in log employment per firm on the change in log real GSP per worker and the change in log lagged number of firms using 1-year changes and cumulative 41-year changes, respectively. We use first differences rather than levels to control for state fixed effects coming from state variation in price-cost markups, the entry cost shifter, etc. We find that average employment per firm does not vary strongly with output per worker, which implies λ^{OLS} in the range of 0.69 to 0.95, depending on the horizon we use and whether we control for lagged number of firms. For the 41-year horizon, which perhaps corresponds the best to our long run framework, the implied λ^{OLS} is 0.94 (s.e. 0.10) when we control for the lagged number of firms. We do not find strong relationship between average employment per firm and lagged firms for state changes.[12]

Figure 2: Across U.S. states, 2020

(a) Number of firms vs. employment

(b) Employment per firm vs. GSP/employment



Notes: The number of firms and employment in each state are from the U.S. Census Bureau's Business Dynamic Statistics (BDS). Real Gross State Product (GSP) is from the U.S. Bureau of Economic Analysis (BEA). In the left panel, the slope coefficient is 0.90 (s.e. 0.02) and the R^2 = 0.97. For the right panel, the slope is -0.14 (0.16) and the R^2 is 0.01.

Table 3 displays the results when we use average *new* firm employment, instead of average employment for all firms, as the dependent variable. The OLS estimates of λ are large and significant, while that for ϕ are larger than when using all firms. Amplification continues to be modest.[13]

[12]Table A3 in Appendix D shows similar results when we run the regression using establishments instead of firms.

[13]See Table A4 in Appendix D for results using new plants.

Table 2: Average firm size on GSP per worker and lagged number of firms

Changes over time, U.S. states, 1978–2020

Horizon	41 years	41 years	1 year	1 year
λ^{OLS}	0.954 (0.07)	0.938 (0.099)	0.694 (0.014)	0.712 (0.014)
ϕ^{OLS}		0.085 (0.062)		-0.047 (0.021)
R^2	0.004	0.043	0.194	0.175
N	100	50	2100	2050
Amplification	1.6% 2.4%	2.3% 3.8%	11.4% 0.6%	10.1% 0.6%

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

4 Inference on λ and ϕ

The previous section shows that average employment per firm is flat or rising in response to output per worker. Through the lens of the free entry condition, this pattern is consistent with entry costs rising with growth. This section considers what values of λ and ϕ could explain why entry costs rise with growth.

As mentioned, our OLS estimates λ^{OLS} and ϕ^{OLS} based on (13) or (14) may be biased because the labor productivity regressor is endogenous to the residual ϵ, which represents demeaned entry efficiency. Higher entry efficiency ϵ should induce more entry and thereby raise labor productivity through the love of variety. If entry efficiency ϵ is independent of the population and

Table 3: *New* firm size on GSP per worker and lagged number of firms

Changes over time, U.S. states, 1978–2020

Horizon	41 years	41 years	1 years	1 year
λ^{OLS}	1.174	1.125	1.154	1.145
	(0.091)	(0.107)	(0.105)	(0.11)
ϕ^{OLS}		0.282		0.028
		(0.067)		(0.161)
R^2	0.036	0.302	0.001	0.001
N	100	50	2100	2050
Amplification	-5.5%	-5.5%	-4.9%	-4.7%
	(2.7%)	(4.4%)	(3.2%)	(3.4%)

See the notes to Table 2.

amenities H_s, however, then we can use these orthogonality conditions to consistently estimate λ and ϕ.[14] See Appendix A for details.

The first row of Table 4 displays our GMM estimates of λ based on national time series, restricting $\phi = 0$. For this case, we only need the single moment condition that A^e is orthogonal to the national population. As the model does not allow $\lambda > 1$, we infer the corner value $\lambda = 1$ (labor intensive entry). To estimate the parameters of the spatial model using cross-state data, we assume entry efficiency A^e_s is orthogonal to 20 year lagged birth rate and amenities H_s across states. The results are presented in the second and third rows of Table 4. We continue to find $\lambda^{GMM} = 1$. The estimated ϕ^{GMM} is -0.14 (s.e. 0.01), which implies a modestly negative knowledge spillover. Such negative spillovers are in the spirit of Bloom et al. (2020).

[14]If there is serial correlation in entry efficiency, then lagged firms could be correlated with the residual. For the time series estimation, we can instrument lagged firms with national population. When we do this the results are similar.

Table 4: Estimated values of λ and ϕ

Model	Assumption	OLS λ	GMM λ	OLS ϕ	GMM ϕ
1. National	$\epsilon_t \perp \ln Pop_t$	1.415 (0.027)	1		
2. Spatial	$\epsilon_{st} \perp$ lagged birth rate$_{st}$	0.954 (0.070)	1		
3. Spatial	$\epsilon_{st} \perp$ lagged birth rate$_{st}$, $\epsilon_{st} \perp \ln H_{st}$	0.938 (0.099)	1	0.085 (0.062)	-0.141 (0.006)

Notes: $\ln Pop_t$ is civilian non-instituionalized population from the Census Bureau and 20 year lagged birth rate is the number of birth per 1000 from the National Center for Health Statistics. $\ln H_s$ is calculated from $\ln L_s - \frac{\alpha}{1-\alpha} \ln \frac{Y_s}{L_s}$ with $\alpha = 0.84$, where Y_s is state real output per worker from the BEA, L_s is state employment from the BDS. For rows 1, λ^{OLS} is from column 1 of Table 1. For rows 2 and 3, λ^{OLS} and ϕ^{OLS} are from columns 1 and 2 of Table 2. λ^{GMM} is restricted to be between 0 and 1. There is no standard error for λ^{GMM} at the upper bound of 1. The standard error of ϕ^{GMM} is the standard error when estimating ϕ^{GMM} while setting $\lambda = 1$.

We now consider the amplification of shocks at our estimated values of λ and ϕ. The first row of Table 5 displays the amplification formula for the real wage response to shocks hitting the growth rates of productivity A, population, or entry efficiency. The second row shows the effect in the special case where $\lambda = 1$ and $\phi = 0$ (entry involves only a fixed amount of labor), under which there is no amplification through entry. The third row considers the special case when $\lambda = 0$ and $\phi = 0$ (entry involves only a fixed amount of output). We find 50% amplification to all three shocks in this special case. If we continue to assume entry costs denominated in output ($\lambda = 0$) but add a *positive* knowledge spillover, then the Table indicates that amplification rises to 200% for productivity shocks and 500% for population or entry efficiency shocks.

The final row of Table 5 calculates amplification under our GMM estimates of λ and ϕ. Since the estimated λ is one and ϕ is negative, we obtain little

Table 5: Ampification of real wage responds to shocks

Parameters	g^A shock	g^L or $\Delta\epsilon$ shocks
General case	$\dfrac{\frac{1-\lambda}{1-\phi}}{\sigma-1-\frac{1-\lambda}{1-\phi}}$	$\dfrac{(\sigma-1)\frac{\phi}{1-\phi}+\frac{1-\lambda}{1-\phi}}{\sigma-1-\frac{1-\lambda}{1-\phi}}$
Special cases		
$\lambda = 1$, $\phi = 0$ (no amplification)	0%	0%
$\lambda = 0$, $\phi = 0$ (λ amplification)	50%	50%
$\lambda = 0$, $\phi = 0.5$ (λ and ϕ amplification)	200%	500%
GMM spatial estimate		
$\lambda = 1$, $\phi = -0.14$	0%	-12%

Notes: Entries show the response of log real wages to a 100% shock to productivity (A), employment (L), or entry efficiency (ϵ) minus the the change when $\lambda = 1$ and $\phi = 0$, expressed as a percent of the change when $\lambda = 1$ and $\phi = 0$. The last row provides the responses using our point estimates for λ and ϕ over time within U.S. states (i.e., our "spatial" estimates). We assume $\sigma = 4$ throughout.

amplification. This suggests that shocks to productivity, population, and entry efficiency are only weakly amplified through induced changes in entry.

5 Empirical robustness checks

In this section we check the robustness of our finding that entry costs rise with growth by considering alternative explanations for the stability of employment per firm with respect to output per worker.

5.1 Discount rate, post-entry growth rate, and exit rate

We used a one-shot model for illustration in the previous section, with no firm life-cycle dynamics. A natural question is whether the stability of employment per firm could reflect changes in the dynamics of firms after entry rather than entry costs rising with growth. Consider an extension of the illustrative model wherein each entrant f in period t draws initial productivity $A_0(f,t)$. After entry, their productivity grows at rate g and they exit at exogenous rate δ. Suppose entrants discount future profits at rate r and that g is sufficiently small relative to δ and r such that the present discounted value of profits is finite. The free entry condition then equalizes the entry cost with the sum of discounted profits. This implies that entry costs are related to the average employment of entrants by

$$\frac{p^e_{st}}{P_{st}} \propto \frac{w_{st}}{P_{st}} \cdot \frac{L^0_{st}}{N^0_{st}} \sum_{a=0}^{\infty} \left(\frac{(1+g)^{\sigma-1}(1-\delta)}{1+r} \right)^a$$

where a is firm age and L^0/N^0 is the average employment of *new* firms. The new-firm results in Tables 1 and 3 say that L^0/N^0 is stable relative to changes in output per worker over time and across states. Our interpretation is that g, r, and δ are stable and entry costs p^e/P rise proportionally with output per worker. An alternative explanation, however, is that entry costs p^e/P are constant but changes in g, r and δ offset the changes in output per worker. For example, if the discount rate r rises with w/P, the post-entry growth rate declines with w/P, or the exit rate rises with w/P this could confound our inference about how entry costs change with growth.

While output per worker rose significantly over time in the U.S., we did not see significantly higher interest rates or return to capital. See, Gomme, Ravikumar and Rupert (2011) and Farhi and François (2018). We do not expect interest rates to vary significantly across states, meanwhile, as capital flows freely across states. Furthermore, studies document that firm exit rate by age has been stable over time while employment growth rate by age has been stable or increasing for older firms — see Karahan, Pugsley and Şahin (2024) and Hopenhayn, Neira and Singhania (2022). This suggests that the present discounted value of profit may have increased even faster with growth than implied by our regressions using new firm employment. Hence, we infer that entry costs rise with growth even after considering post-entry dynamics.

For U.S. manufacturing we can go a step further and directly calculate the present discounted value of profits. For a cohort of entering in year t, we calculate the expected PDV of profits using the average *realized* PDV of profits for a cohort. So our proxy for entry costs in period t is

$$\frac{1}{N_e(t)} \sum_{f=1}^{N_e(t)} \sum_{a=0}^{D_f} \beta(t,a) \cdot \pi_f(t,a) \tag{15}$$

where $N_e(t)$ is the number of entrants, f indexes the entrants in the cohort, and D_f denotes the age of the entrant at the time of exit (death). Here $\pi_f(t,a)$ is the profit of entrant f from cohort t and age a and $\beta(t,a)$ is the discount factor.

Implementing the PDV measure requires us to estimate the flow of profits. Rather than trying to distinguish economic and accounting profits or variable and fixed costs, we estimate price-cost markups and combine our estimates with revenue to infer profits. Although estimating the *level* of markups is notoriously difficult, for our purposes we only need to know how markups vary over time. We follow Bils, Klenow and Malin (2018) in using the inverse ratio of shipments to the sum of payroll costs and intermediate inputs expenditures.[15]

We use establishment-level data from the U.S. Census Bureau's Census of Manufacturing (CMF) for 1963 and quinquennially from 1972 to 2012. The CMF covers all establishments with employees. For our sample period, there are about 1.54 million unique establishments. We construct cohorts based on the first year each establishment appears in the data. This means that we drop all observations in 1963, because we cannot identify when these plants

[15]Using the sum of payroll and intermediate inputs addresses concerns with outsourcing trends raised by Giannoni and Mertens (2019) and others.

entered; we use the 1963 plants to determine which of the 1967 plants are entrants. We also drop about 7,600 plants that exit and then re-enter, as their entry year is ambiguous. We drop all plant-years with negative or missing shipments, intermediate inputs, payroll or employment.

We calculate the PDV of profits for each cohort in the following way. First, we multiply shipments by the profit share (implied by our time-varying inverse shipment to payroll and intermediate costs) to generate profits for each plant-year. We deflate all profits by the BEA manufacturing value added deflator.[16] We discount each year of real profits assuming a constant real interest rate $r = 0.05$. We calculate the PDV of real profits for each cohort using horizons of 0, 5, 10 and 15 years. A shorter horizon gives us more observations but covers less of a cohort's lifetime. The PDV for each cohort should be an unbiased estimate of its entry cost, given the zero profit condition for entrants.

We use real manufacturing value added per worker each year to proxy for aggregate productivity. We deflate total value added per worker in each year by the BEA manufacturing value added deflator. We calculate the total value added and total number of workers by summing value added and employment across plants in each year.

Across entering cohorts, we regress the log of the PDV of real profits on the log of real value added per worker in the year of the cohort's entry. Table 6 presents the results. At the 5 and 10 year horizons, the PDV of profits rises even more than one-for-one with labor productivity at the time of entry (a slope above 1). The standard errors are small (0.25 or less) and the R^2's are large (0.8 or higher). At the 15-year horizon the PDV of profits increases less than one-for-one with labor productivity at entry, but the connection is still quite positive (slope 0.65). Thus, at all horizons, it appears that entry costs rise strongly with average labor productivity in U.S. manufacturing.

[16]Since we only have data every five years, for each plant we interpolate real profit between years to generate yearly profits. We linearly interpolate the log of real profits, which is equivalent to fitting a constant growth rate of real profits between adjacent observations.

Table 6: PDV of establishment profits on value added per worker

U.S. manufacturing; 1967, 1972,..., 2012

Horizon	0	5	10	15
Coefficient on $\ln Y/L$	1.297	1.232	1.254	0.648
	(0.097)	(0.191)	(0.250)	(0.151)
R^2	0.957	0.856	0.807	0.787
# of cohorts	10	9	8	7
First cohort	1967	1967	1967	1967
Last cohort	2012	2007	2002	1997

Note: U.S. Census of Manufacturing (CMF) and the U.S. Bureau of Economic Analysis (BEA). The table reports the regression coefficient from regressing log real PDV of profits by cohort on log real manufacturing output per worker at the time of entry. Horizon h means the PDV is calculated using profit streams from age 0 to age h.

5.2 Trends in the aggregate markup or markdown

The relationship between entry costs, average employment per firm, and the real wages is

$$\frac{p^e_{st}}{P_{st}} = \frac{1}{\sigma_{st}-1}\left(\frac{wL}{PN}\right)_{st}$$

Our baseline model assumes the elasticity of substitution σ_{st} is either constant over time or homogenous across states. Thus an alternative explanation for the stability of firm employment with respect to output per worker is that markups are declining (say due to σ rising) with output per worker. Similarly, extending our model to include time-varying *markdowns* would imply

$$\frac{p^e_{st}}{P_{st}} = \frac{1}{\eta_{st}}\left(\frac{wL}{PN}\right)_{st}$$

where η_{st} is the elasticity of labor supply. A lower η_{st} implies more labor market power for firms and a steeper markdown.

Looking at the above expressions, we see that entry costs may be constant ($\lambda = 0$) even if average firm employment is constant because firm product and labor market power decline proportionally with output per worker over time or across states, such that $\frac{1}{\sigma-1}\frac{w}{P}$ or $\frac{1}{\eta}\frac{w}{P}$ in the above expressions is constant over time and across states. Intuitively, when entry costs are stable with respect to changes in output per worker, higher labor productivity reduces equilibrium average employment per firm. In the opposite direction, weaker product and labor market power (higher σ and η) raise employment per firm because more revenue is needed to generate the same amount of profits. In theory, equilibrium firm employment may not vary with output per worker because these two forces exactly cancel each other out.

Our regression of within-state changes in (14) controls for markup and markdown heterogeneity across states that can be picked up by state fixed effects — i.e., σ and η variation across states but not over time. Over time in the U.S., the PDV calculations in the previous section control for markup and markdown trends, at least for manufacturing. We also ran (14) with a time fixed effect to control for changes in markups and markdowns over time. Note that we cannot run this regression for the longest horizon in Table 7 because we only have one period in that case. For the 10 year and 1 year horizons, we find similar coefficients to our baseline regression in Table 2.

In addition to these robustness checks, it is worth noting that the literature tends to find rising or stable markups. See, for example, Autor, Dorn, Katz, Patterson and Van Reenen (2020) and De Loecker, Eeckhout and Unger (2020). Evidence for markdowns is more mixed — Berger, Herkenhoff and Mongey (2022) report a decline in local labor market concentration between 1977 and 2013, while Yeh, Macaluso and Hershbein (2022) find that markdowns declined and then increased over time.

Table 7: Average firm size on GSP per worker and lagged number of firms

Changes over time, U.S. states, 1978–2020, with time fixed effects

Horizon	10 years	10 years	1 year	1 year
λ^{OLS}	0.626	0.634	0.688	0.708
	(0.048)	(0.061)	(0.014)	(0.014)
ϕ^{OLS}		0.016		-0.066
		(0.079)		(0.022)
within R^2	0.385	0.385	0.201	0.183
N	150	150	2100	2100
Amplification	14.2%	14.2%	11.6%	10.0%
	2.1%	2.1%	0.6%	0.6%

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

5.3 Industry composition

Our inference is based on a single industry model. We can easily extend the inference to multiple industries. Suppose industry output is produced using the CES structure as in our baseline model. And suppose entry into an industry uses c_i units of the entry good. Then free entry into each industry implies that average employment in an industry is equal to $(\sigma - 1 + \lambda)c_i \frac{p^e}{w}$. Aggregate employment per firm is then the industry-weighted average of entry costs relative to wages:

$$\frac{L}{N} = (\sigma - 1 + \lambda)\frac{p^e \sum_i (\frac{N_i}{N} c_i)}{w}.$$

Therefore, the empirical pattern of stable L/N relative to output-per-worker still implies that entry costs $p^e \sum_i (\frac{N_i}{N} c_i)$ rise with growth. However, in addition to the entry technology channels (ϕ and λ) that work through p^e, the rise in entry costs can also be explained by reallocation towards industries with higher entry costs c_i. We can distinguish between the reallocation and entry technology channels by using a measure of average employment that is not affected by reallocation across industries. Let $\bar{s}_i$ be the average share of firms in industry i across years. If the free entry condition holds in each industry then

$$\sum_i \bar{s}_i \frac{L_i}{N_i} = (\sigma - 1 + \lambda) \frac{p^e}{w} \left(\sum_i \bar{s}_i c_i \right).$$

Changes in this fixed-weight average come purely from changes in p^e/w

$$d \ln \left(\sum_i s_i \frac{L_i}{N_i} \right) = d \ln \left(\frac{p^e}{w} \right).$$

The BDS data reports employment and firms by NAICS 2-digit in each state-year for the nonfarm business sector. We set s_i in each state to the 1978–2020 average NAICS 2-digit share of firms. Table 8 shows the same regression as Table 2 but with the fixed-weight on each industry in constructing average employment per firm on the left hand side. The results are similar to the baseline in Table 2.

5.4 Selection on entry

Our inference strategy assumes the entrants do not know their productivity before entering and hence entry costs are proportional to average firm employment. If entrants know their productivity, however, then the free entry condition implies that entry costs are proportional to the employment of the *marginal* entrant rather than to average employment across all entrants.

In the case of Pareto-distributed entrant productivity, the size of the

Table 8: Average firm size on GSP per worker and lagged number of firms

Changes over time, U.S. states, 1978–2020, fixed industry weights

Horizon	41 years	41 years	1 year	1 year
λ^{OLS}	1.013	0.988	0.689	0.721
	(0.072)	(0.102)	(0.015)	(0.015)
ϕ^{OLS}		0.080		-0.130
		(0.064)		(0.022)
R^2	0.000	0.032	0.179	0.168
N	100	50	2100	2000
Amplification	-0.4%	0.4%	11.6%	9.0%
	(2.4%)	(3.7%)	(0.6%)	(0.6%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

marginal entrant is proportional to that of the average entrant. In this case our finding that average entrant size increases with labor productivity is consistent with entry costs rising with growth. In the event of normally distributed entrant productivity, however, we need to examine other moments of the entrant size distribution. In particular, if the profit of the marginal entrant is pinned down by a constant entry cost denominated in terms of output, then we expect the dispersion of profits to increase with output per worker under normally distributed entrant productivity.

We can look at entrant dispersion in U.S. manufacturing over time. Table 9 displays the results from regressing dispersion in PDV against real output per worker in U.S. manufacturing. At all horizons, dispersion fails to increase with output per worker. Hence, we conclude that entry costs faced by the marginal entrant must also be increasing with output per worker.[17]

Table 9: Dispersion of PDV of profits of new estab. on value added per worker

U.S. manufacturing; 1967, 1972,..., 2012

Horizon	0	5	10	15
Coefficient on Y/L	-0.091	-0.099	-0.018	-0.374
	(0.068)	(0.116)	(0.157)	(0.097)
R^2	0.180	0.093	0.002	0.748
# of cohorts	10	9	8	7
First cohort	1967	1967	1967	1967
Last cohort	2012	2007	2002	1997

Note: U.S. Census of Manufacturing (CMF) and the U.S. Bureau of Economic Analysis (BEA). The table reports the regression coefficient from regressing dispersion of log real PDV of profits by cohort on log real manufacturing output per worker at the time of entry. Horizon h means the PDV is calculated using profit streams from age 0 to age h.

5.5 Measurement error in labor

The modest relationship we find between average employment per firm and labor productivity across time and states could be biased downward by measurement error in labor L. We check whether our results are driven by this division bias using employment from the County Business Patterns (CBP) to construct gross state product per worker but employment from the Business Dynamics Statistics to construct employment per firm. Table 10 displays the results.[18] The regression coefficients are similar to the baseline in Table 2. We

[17] Appendix Table A5 shows that average employment of firms with 1-4 employees is also stable with respect to output per worker across states. This is consistent with entry costs of marginal entrants rising with growth if the marginal entrants are in the smallest employee bin.

[18] The longest horizon is shorter than in the baseline due to availability of CBP data.

also checked for potential measurement errors by using Y/L^{CBP} to instrument for Y/L^{BDS}. The results are similarly reassuring, and are reported in Appendix Tables A6. Thus, our results do not appear to be from measurement error in L.

Table 10: BDS workers per firm on GSP per CBP worker and lagged # of firms

Changes over time, U.S. states, 1986–2020

Horizon	34 years	34 years	1 year	1 year
λ^{OLS}	0.956	0.950	0.801	0.803
	(0.101)	(0.098)	(0.015)	(0.015)
ϕ^{OLS}		0.137		-0.036
		(0.069)		(0.023)
R^2	0.004	0.082	0.095	0.097
N	50	50	1700	1700
Amplification	1.5%	2.0%	7.1%	6.7%
	(3.5%)	(3.9%)	(0.6%)	(0.6%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau and the County Business Patterns. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

6 Conclusion

In the U.S., the number of worker per firm is stable or rises with output per worker over time and across states. This fact 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. This may involve a negative knowledge spillover from past innovation *a la* Bloom et al. (2020). 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 boost entry of firms. Thus there is no amplification of the effect on aggregate productivity through entry. Second, if the modelling choice is between fixing entry costs in labor or output, it is more realistic to denominate 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.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr, "Innovation, Reallocation, and Growth," *American Economic Review*, 2018, *108* (11), 3450–91.

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

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

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

_ and _ , "Aggregate Implications of Innovation Policy," *Journal of Political Economy*, 2019, *127* (6), 2625–2683.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Baqaee, David and Emmanuel Farhi, “Entry vs. Rents: Aggregation with Economies of Scale,” Technical Report, UCLA working paper 2021.

Baqaee, David Rezza, Emmanuel Farhi, and Kunal Sangani, “The Darwinian Returns to Scale,” *Review of Economic Studies*, 2023.

Bento, Pedro and Diego Restuccia, “Misallocation, Establishment Size, and Productivity,” *American Economic Journal: Macroeconomics*, July 2017, *9* (3), 267–303.

_ and _ , “The Role of Nonemployers in Business Dynamism and Aggregate Productivity,” *NBER working paper 25998*, 2022.

Berger, David, Kyle Herkenhoff, and Simon Mongey, “Labor Market Power,” *American Economic Review*, 2022, *112* (4), 1147–93.

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.

Berry, Steven and Joel Waldfogel, “Product quality and market size,” *The Journal of Industrial Economics*, 2010, *58* (1), 1–31.

Bhattacharya, Dhritiman, Nezih Guner, and Gustavo Ventura, “Distortions, Endogenous Managerial Skills and Productivity Differences,” *Review of Economic Dynamics*, 2013, *16* (1), 11–25.

Bilbiie, Florin, Fabio Ghironi, and Marc Melitz, “Endogenous Entry, Product Variety and Business Cycles,” *Journal of Political Economy*, 2012, *120* (2), 304–345.

Bils, Mark, Peter J Klenow, and Benjamin A Malin, “Resurrecting the Role of the Product Market Wedge in Recessions,” *American Economic Review*, 2018, *108* (4-5), 1118–1146.

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.

— , Charles I Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, 2020, *110* (4), 1104–1144.

Boar, Corina and Virgiliu Midrigan, "Efficient Redistribution," *Journal of Monetary Economics*, 2022, *131*, 78–91.

— and — , "Markups and Inequality," Technical Report 2022.

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.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the US Economy," *The Review of Economic Studies*, 2018, *85* (4), 2042–2096.

Clementi, Gian Luca and Dino Palazzo, "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, 2016, *8*, 1–41.

Cole, Harold L, Jeremy Greenwood, and Juan M Sanchez, "Why Doesn't Technology Flow from Rich to Poor Countries?," *Econometrica*, 2016, *84* (4), 1477–1521.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

David, Joel M, "The Aggregate Implications of Mergers and Acquisitions," *The Review of Economic Studies*, 2021, *88* (4), 1796–1830.

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.

Farhi, Emmanuel and Gourio François, "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia," *Brookings Papers on Economic Activity*, 2018, *2018* (2), 147–250.

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.

Fujimoto, Takao and Ulrich Krause, “Strong Ergodicity for Strictly Increasing Nonlinear Operators,” *Linear Algebra and its Applications*, 1985, *71*, 101–112.

Giannoni, Marc and Karel Mertens, “Outsourcing, Markups and the Labor Share,” *Federal Reserve Bank of Dallas and CEPR Working Paper*, 2019.

Gomme, Paul, B. Ravikumar, and Peter Rupert, “The Return to Capital and the Business Cycle,” *Review of Economic Dynamics*, 2011, *14* (2), 262–278.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, “Long-run Implications of Investment-specific Technological Change,” *The American economic review*, 1997, pp. 342–362.

Grossman, Gene M and Elhanan Helpman, “Quality Ladders in the Theory of Growth,” *The Review of Economic Studies*, 1991, *58* (1), 43–61.

Gutiérrez, Germán and Thomas Philippon, “The Failure of Free Entry,” Technical Report, National Bureau of Economic Research 2019.

Gutierrez, German, Callum Jones, and Thomas Philippon, “Entry Costs and the Macroeconomy,” Working Paper 25609, National Bureau of Economic Research February 2019. Series: Working Paper Series.

Gutiérrez, Germán, Callum Jones, and Thomas Philippon, “Entry Costs and Aggregate Dynamics,” *Journal of Monetary Economics*, 2021, *124*, S77–S91.

Hopenhayn, Hugo A, “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 1992, pp. 1127–1150.

Hopenhayn, Hugo and Richard Rogerson, “Job Turnover and Policy Evaluation: A General Equilibrium Analysis,” *Journal of Political Economy*, 1993, *101* (5), 915–938.

— , Julian Neira, and Rish Singhania, "From Population Growth to Firm Demographics: Implications for Concentration, Entrepreneurship and the Labor Share," *Econometrica*, 2022, *90* (4), 1879–1914.

Jones, Charles I, "R & D-based Models of Economic Growth," *Journal of political Economy*, 1995, *103* (4), 759–784.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin, "How Free is Free Entry," Technical Report, work in progress 2023.

— , — , and — , "Demographic Origins of the Startup Deficit," *Forthcoming in the American Economic Review*, 2024.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Laincz, Christopher A and Pietro F Peretto, "Scale Effects in Endogenous Growth Theory: An Error of Aggregation Not Specification," *Journal of Economic Growth*, 2006, *11* (3), 263–288.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

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.

— , "Models of Growth and Firm Heterogeneity," *Annu. Rev. Econ.*, 2010, *2* (1), 547–576.

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

Peters, Michael and Conor Walsh, "Population Growth and Firm Dynamics," *Working Paper*, 2022.

Redding, Stephen J, "Trade and Geography," *Handbook of International Economics*, 2022, *5*, 147–217.

— and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9*, 21–58.

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.

Sterk, Vincent, Petr Sedlek, and Benjamin Pugsley, "The Nature of Firm Growth," *American Economic Review*, 2021, *111* (2), 547–79.

Walsh, Conor, "The Startup Multiplier," *working paper*, 2023.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein, "Monopsony in the US Labor Market," *American Economic Review*, July 2022, *112* (7), 2099–2138.

A Derivations for models

A.1 Derivations for the love-for-variety model

This section provides derivation for the national love-for-variety model in Section 2.1 and the estimation of the model in Section 4.

Endowment: L units of labor

Technology: A, A^e are exogenous.

$$\begin{aligned} y_i &= Al_i \quad \text{(Intermediate goods production)} \\ Y &= \left[\int^N y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} \quad \text{(Final goods production)} \\ N &= A^e (L^e)^{\lambda} (Y^e)^{1-\lambda} \quad \text{(Entry goods production)} \end{aligned}$$

Numeraire: Final goods is the numeraire $P = 1$

Firm's problem:

$$\begin{aligned} &\max_{\{y_i\}_i} \quad Y - \int_i p_i y_i \, di, \quad s.t. \quad Y \le \left[\int^N y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} \quad \text{(Final goods producer)} \\ &\pi_i \equiv \max_{y_i, l_i} \quad p_i y_i - w l_i, \quad s.t. \quad y_i \le A l_i, \quad y_i = p_i^{-\sigma} Y \quad \text{(Intermediate goods producer)} \\ &\max_{Y_e, L_e} \quad p^e N - Y^e - w L^e, \quad s.t. \quad N \le A^e (L^e)^{\lambda} (Y^e)^{1-\lambda} \quad \text{(Entry goods producer)} \end{aligned}$$

Zero-profit-condition:

$$N(p^e - \pi_i) = 0, \quad N \ge 0, \quad \pi_i \ge p^e$$

Household's problem:

$$\max_{C} \quad u\left(\frac{C}{L}\right), \quad s.t. \quad C \le wL + \pi N - p^e N$$

Market clearing conditions:

$$L = L^e + L^y, \quad Y = C + Y^e$$

Solving the intermediate goods producer's problem yields

$$p_i = \frac{w}{A}\frac{\sigma}{\sigma-1}, \quad \pi_i = \frac{wl_i}{\sigma-1} = \frac{p_i y_i}{\sigma}$$

while solving the entry goods producer's problem yields

$$p^e = \frac{1}{A_e}\left(\frac{w}{\lambda}\right)^{\lambda}\left(\frac{1}{1-\lambda}\right)^{1-\lambda}$$

Combining these solutions with the labor market clearing condition and the zero-profit-condition, we derive the share of total labor in producing intermediate goods

$$\frac{L^y}{L} = \frac{wL^y}{wL^y + wL^e} = \frac{(\sigma-1)\pi_i}{(\sigma-1)\pi_i + \lambda\pi_i} = \frac{\sigma-1}{\sigma-1+\lambda}.$$

As a corollary,

$$\frac{L^e}{L} = \frac{\lambda}{\sigma-1+\lambda},$$

Substituting the solutions for L^y into the final goods production function, the relationship between π_i and wl_i, the entry goods production function and price of entry goods, we get the following relationships

$$\begin{aligned}
Y &= AL^y N^{\frac{1}{\sigma-1}} = \frac{\sigma-1}{\sigma-1+\lambda} ALN^{\frac{1}{\sigma-1}} \\
w &= \frac{\sigma-1}{\sigma}\frac{Np_i y_i}{L^y} = \frac{\sigma-1+\lambda}{\sigma}\frac{Y}{L} \\
Np^e &= \frac{wL^e}{\lambda} = \frac{wL}{\sigma-1+\lambda} \\
p^e &= \frac{1}{A^e}\left(\frac{w}{\lambda}\right)^{\lambda}\left(\frac{1}{1-\lambda}\right)^{1-\lambda}
\end{aligned}$$

Combining these equations and expressing in natural logs, we have the equation that links employment per firm with output per worker

$$\ln\frac{L}{N} = \text{constant} - (1-\lambda)\ln\frac{Y}{L} - \ln(A^e)$$

where constant $= \lambda\ln(\sigma-1+\lambda) + (1-\lambda)\ln\sigma - \lambda\ln(\lambda) - (1-\lambda)\ln(1-\lambda)$.

Furthermore, we can derive the following simultaneous equations that relates w, N and p^e to the exogenous variables.

$$\begin{aligned}
&\ln N + \ln\frac{p^e}{w} = \ln L - \ln(\sigma-1+\lambda) =: b_{pop}\\
&\ln w - \frac{1}{\sigma-1}\ln N = \ln\frac{\sigma-1}{\sigma} + \ln A =: b_{tech}\\
&\lambda\ln w - \ln p^e = \ln A^e + \lambda\ln\lambda + (1-\lambda)\ln(1-\lambda) =: b_{entry}
\end{aligned}$$

Solving these gives the following equations for the endogenous variables in terms of the exogenous variables.

$$\begin{aligned}
\ln w &= \frac{(\sigma-1)b_{tech} + b_{pop} + b_{entry}}{\sigma-1-(1-\lambda)}\\
\ln N &= (\sigma-1)(\ln w - b_{tech}) = (\sigma-1)\frac{(1-\lambda)b_{tech} + b_{pop} + b_{entry}}{\sigma-1-(1-\lambda)}\\
\ln p^e &= \lambda\ln w - b_{entry} = \frac{\lambda(\sigma-1)b_{tech} + \lambda b_{pop} - (\sigma-2)b_{entry}}{\sigma-1-(1-\lambda)}
\end{aligned}$$

The welfare analysis in Section 2.1 follows directly from these three equations. The first equation gives the welfare impact of changes in productivity, population, and entry efficiency. The second equation illustrates the variety expansion channel. The number of varieties N responds to changes in production productivity only if the goods share of entry is positive. Finally, the last equation shows the entry costs rise with exogenous productivity and population only if the labor share of entry is positive.

A.1.1 GMM

Next we detail the procedure used for the estimation in Section 4 pertaining to the national model. As in the main text, we introduce spillover from past entry in the form of $A_e = e^{\epsilon} N_{-1}^{\phi}$, where N_{-1} is the number of firms from the previous period. Substituting this into the regression equation above

$$\ln \frac{L_t}{N_t} = \xi - (1-\lambda) \ln \frac{Y_t}{L_t} - \phi \ln(N_{t-1}) - \epsilon_t$$

where $\xi = \lambda \ln(\sigma - 1 + \lambda) + (1-\lambda) \ln \sigma - \lambda \ln(\lambda) - (1-\lambda) \ln(1-\lambda)$. Since Y/L is endogenous to ϵ, the OLS estimates of λ and ϕ are biased. In the GMM inference, we assume ϵ is independent of productivity which implies $Cov(\epsilon_t, \ln A_t) = 0$ to estimate λ when we restrict $\phi = 0$. We additionally impose independence between population Pop_t and ϵ_t which implies $Cov(\epsilon_t, \ln Pop_t) = 0$ when we jointly estimate λ and ϕ.

Population is observable while, according to the model, productivity can be measured from the data by

$$\ln A_t = \ln \frac{Y}{L}_t - \frac{1}{\sigma - 1} \ln N_t + \ln \frac{\sigma - 1 + \lambda}{\sigma - 1}.$$

Therefore, we have the following identifying restrictions for λ and ϕ

$$\mathbb{E}[g_t] = 0, \quad g_t := \begin{bmatrix} g_t^1 \\ g_t^2 \end{bmatrix} := \begin{bmatrix} \widetilde{\epsilon}_t \widetilde{\ln Pop_t} \\ \widetilde{\epsilon}_t \widetilde{\ln A_t} \end{bmatrix}$$

where

$$-\widetilde{\epsilon}_t = \widetilde{\ln\left(\frac{L}{N}\right)} + (1-\lambda) \cdot \widetilde{\ln\left(\frac{Y}{L}\right)} + \phi \cdot \widetilde{\ln(N_{-1})}$$

The tilde notation denotes the deviation of a variable from its expected value. We construct the sample analogue of ϵ by using the deviation from the sample mean for $\ln \frac{L}{N}$, $\ln \frac{Y}{N}$ and $\ln N_{-1}$. As in the main text, we set $\sigma = 4$. When we impose $\phi = 0$, our moment condition is $\mathbb{E}[g_t^2] = 0$.

A.2 The spatial equilibrium model

This appendix describes the solution to the spatial model in Section 2. First, from the government budget constraint and the zero profit condition of the firms, the income of each worker in a state satisfies

$$v_s = w_s + (1-\alpha)v_s = \frac{w_s}{\alpha}.$$

Hence housing demand per worker is higher when wages increase relative to rent

$$h_s = \frac{(1-\alpha)v_s}{r_s} = \frac{w_s}{r_s}\frac{1-\alpha}{\alpha}.$$

Substituting this relationship into the land market clearing condition pins down rent r_s given H_s, L_s and w_s

$$r_s = \frac{w_s}{\alpha}\frac{1-\alpha}{h_s}, \quad h_s = \frac{H_s}{L_s}.$$

Furthermore, since the marginal cost of a unit of utility in each location is $P_s^{\alpha} r_s^{1-\alpha}$, workers being indifferent between states implies that there exists $\bar{V}$ such that

$$\frac{v_s}{P_s^{\alpha} r_s^{1-\alpha}} = \bar{V} \quad \forall s. \tag{A1}$$

Combining this condition with the relationship between r_s and v_s above, we can derive the following expression for welfare

$$\bar{V} = \left(\frac{H_s}{L_s}\right)^{1-\alpha} \left(\frac{w_s}{P_s}\right)^{\alpha} \left(\frac{1}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\alpha}\right)^{\alpha}. \tag{A2}$$

This says that areas with higher wages have smaller (quality-adjusted) dwellings per worker. Since population across states must sum to the exogenous total population L, we have

$$L = \sum_{s=1}^{S} H_s \left\{ \frac{1}{\bar{V}} \left(\frac{w_s}{P_s}\right)^{\alpha} \left(\frac{1}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\alpha}\right)^{\alpha} \right\}^{1/(1-\alpha)}.$$

This solves for $\bar{V}$ given real wages across states

$$\bar{V} = \left(\frac{1}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\alpha}\right)^{\alpha} \left\{ \sum_{s=1}^{S} \frac{H_s}{L} \left(\frac{w_s}{P_s}\right)^{\alpha/(1-\alpha)} \right\}^{1-\alpha}. \tag{A3}$$

We follow the method of Allen and Arkolakis (2014) to solve for real wages. First, rearranging (A2), we have

$$\frac{w_s}{P_s} \left(\frac{H_s}{L_s}\right)^{\frac{1-\alpha}{\alpha}} = \overline{V}^{\frac{1}{\alpha}} \alpha (1-\alpha)^{\frac{1-\alpha}{\alpha}} \equiv \overline{W}. \tag{A4}$$

Then substituting in the expression for P_s in terms of equilibrium $p_{s,s'}$ and $N_{s'}$, we arrive at

$$w_s \left(\frac{H_s}{L_s}\right)^{\frac{1-\alpha}{\alpha}} = \overline{W} \left(\sum_{s'} N_{s'} p_{s,s'}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{A5}$$

The number of firms N_s in turn is related to the population in the state s through the free entry and labor market clearing conditions

$$L_s = L_s^e + L_s^y = \frac{\lambda N_s p_s^e}{w_s} + (\sigma - 1)\frac{N_s p_s^e}{w_s} = (\sigma - 1 + \lambda)\frac{N_s p_s^e}{w_s}.$$

Substitute this and $p_{s,s'}$ into (A5), we have

$$w_s \left(\frac{H_s}{L_s}\right)^{\frac{1-\alpha}{\alpha}} = \overline{W} \frac{\sigma}{\sigma - 1} \left(\frac{1}{\sigma - 1 + \lambda} \sum_{s'} L_{s'} \frac{w_{s'}}{p_{s'}^e} \left(\frac{w_{s'} d_{s,s'}}{A_{s'}}\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{A6}$$

Following Allen and Arkolakis (2014), we can show that there exists ζ such that equilibrium wage satisfies

$$\zeta = w_s^{1-2\sigma} \frac{w_s}{p_s^e} A_s^{\sigma-1} \left(\frac{H_s}{L_s}\right)^{\frac{1-\alpha}{\alpha}(1-\sigma)}. \tag{A7}$$

Substituting the wage function (A7) with ζ normalized to 1, entry cost function

(7) and (A4) into (A6) yields

$$\left(A_s^e\lambda^\lambda(1-\lambda)^{1-\lambda}\right)^{\frac{1-\sigma}{2\sigma-1}} A_s^{-\frac{(1-\sigma)^2}{2\sigma-1}} \left(\frac{H_s}{L_s}\right)^{\frac{1-\alpha}{\alpha}\frac{(\sigma-1+\lambda)(1-\sigma)}{2\sigma-1}} \left(\frac{\sigma}{\sigma-1}\right)^{\sigma-1}(\sigma-1+\lambda)$$
$$= \overline{W}^{1-\sigma+1-\lambda}\sum_{s'} L_{s'}\left(A_{s'}^e\lambda^\lambda(1-\lambda)^{1-\lambda}\right)^{\frac{\sigma}{2\sigma-1}} d_{s,s'}^{1-\sigma}\left(\frac{H_{s'}}{L_{s'}}\right)^{\frac{1-\alpha}{\alpha(2\sigma-1)}((1-\sigma)^2-\sigma(1-\lambda))} A_{s'}^{\frac{(\sigma-1)\sigma}{2\sigma-1}}. \tag{A8}$$

The exponent on L_s in the left hand side can be written as $\widetilde{\sigma}\gamma_1$ where $\widetilde{\sigma} = \frac{\sigma-1}{2\sigma-1}$ and $\gamma_1 = \frac{1-\alpha}{\alpha}(\sigma-1+\lambda)$. The exponent on the right hand side can be expressed as $\widetilde{\sigma}\gamma_1\frac{\gamma_2}{\gamma_1}$ where $\gamma_2 = 1+\frac{\sigma}{\sigma-1}+\left(\frac{\sigma(1-\lambda)}{\sigma-1}-(\sigma-1)\right)\frac{1-\alpha}{\alpha}$. Applying Fujimoto and Krause (1985), one can shown that as long as $\frac{\gamma_2}{\gamma_1}\in(0,1)$, iterating on (A8) from any initial $\{L_s^0\}_s$ will converge to the equilibrium $\{L_s^*\}_s$. More precisely, from (A8), let T denote operator

$$T(\{L_s\}) = \left(\frac{\overline{W}^{1-\sigma+1-\lambda}\sum_{s'}(A_{s'}^e)^{\frac{\sigma}{2\sigma-1}} d_{s,s'}^{1-\sigma}(H_{s'})^{\frac{(1-\alpha)((1-\sigma)^2-\sigma(1-\lambda))}{\alpha(2\sigma-1)}} L_{s'}^{\widetilde{\sigma}\gamma_1\frac{\gamma_2}{\gamma_1}} A_{s'}^{\frac{(\sigma-1)\sigma}{2\sigma-1}}}{(A_s^e)^{\frac{1-\sigma}{2\sigma-1}} A_s^{-\frac{(1-\sigma)^2}{2\sigma-1}} (H_s)^{-\widetilde{\sigma}\gamma_1}\left(\frac{\sigma}{\sigma-1}\right)^{\sigma-1}\frac{\sigma-1+\lambda}{\lambda^\lambda(1-\lambda)^{1-\lambda}}}\right)^{\frac{1}{\widetilde{\sigma}\gamma_1}}.$$

For any $\{L_s^0\}\neq 0$ and $\frac{T^k\{L_s^0\}}{|T^k\{L_s^0\}|}\longrightarrow\{\widetilde{L}_s^*\}$ and $L_s^* = L\frac{\widetilde{L}_s^*}{\sum\widetilde{L}_s^*}$.

A.2.1 GMM

We derived the following relationships in the main text.

$$\ln\frac{L_s}{N_s} = \text{constant} + (\lambda-1)\ln\frac{GSP_s}{L_s} - \phi\ln N_{s,t-1} - \epsilon_s. \tag{A9}$$

$$\Delta\ln\frac{L_{s,t}}{N_{s,t}} = (\lambda-1)\Delta\ln\frac{GSP_{s,t}}{L_{s,t}} - \phi\Delta\ln N_{s,t-1} - \Delta\epsilon_{st}. \tag{A10}$$

Unlike the national model, the state population is endogeneous to ϵ_s and hence cannot be used as an instrument for GSP_s/L_s. Instead, we assume that within each period t, $\epsilon_{s,t}$ is independent of state amenities $H_{s,t}$ and state 20 year lagged birth rate. More specifically, we use $\mathbb{E}[(\epsilon_{s,t}-\mathbb{E}\epsilon_{s,t})\ln birth_{s,t}] = 0$ for each period t

to estimate λ when we restrict $\phi = 0$ and additionally use $\mathbb{E}[(\epsilon_{s,t} - \mathbb{E}\epsilon_{s,t}) \ln H_{s,t}] = 0$ for each t when jointly estimating λ and ϕ.

For $H_{s,t}$, we use model equation (A4) which implies that which each period t

$$\ln H_s \;=\; \text{constant} \;-\; \frac{\alpha}{1-\alpha} \ln \frac{w_s}{P_s} \;+\; \ln L_s$$

As in the main text, we set $\sigma = 4$. Greenwood et al. (1997) find the share of labour, structures, and equipment, in value added for the U.S. economy are 70%, 13%, and 17% respectively. As our model does not have equipment, we set $\alpha = 70/(70+13)$.[19]

Therefore, we have the following identifying restrictions for λ and ϕ

$$\mathbb{E}g_{s,t} = 0, \quad g_{s,t} := \left[\begin{array}{c} g^1_{s,t} \\ g^2_{s,t} \end{array} \right] := \left[\begin{array}{c} \widetilde{\epsilon_{s,t}}\widetilde{\ln H_{s,t}} \\ \widetilde{\epsilon_{s,t}}\ln \widetilde{birth}_{s,t-20} \end{array} \right]$$

where

$$-\widetilde{\epsilon_{s,t}} = \ln \widetilde{\left(\frac{L}{N}\right)}_{s,t} + (1-\lambda)\ln \widetilde{\left(\frac{Y}{L}\right)}_{s,t} + \phi\widetilde{\ln(H_{s,t})}$$

The tilde notation denotes the deviation of a variable from the average across states.

[19] Caliendo et al. (2018) uses a similar methodology to assign a value to α.

B Welfare and entry costs in other models

In the main text, we showed that entry costs rising with growth matters in the love-of-variety model. In this section, we show that it matters for welfare in several other models as well.

B.1 Static span-of-control model

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

$$Y = \sum_{i=1}^{N} y_i$$

$$y_i = A l_i^{\gamma}$$

$$N = A^e Y_e^{1-\lambda} L_e^{\lambda}$$

The first equation says aggregate output is the simple sum of firm output levels. The second equation specifies the diminishing returns to production technology for each firm ($\gamma < 1$). The third equation is the entry technology. 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, Eifert, Mahajan, McKenzie and Roberts (2013) for example, argue that overhead costs include some information technology equipment.

In the equilibrium, real wage is given by

$$\ln w = \text{constant} + \frac{\frac{1}{1-\gamma} \ln A + \ln A^e}{\frac{1}{1-\gamma} - (1-\lambda)}$$

which increases with production and entry efficiency. The welfare impact of a change in A 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.

B.2 Static love-of-variety model with congestion

Consider the static version of our baseline model but with only one region. Suppose that the entry technology now depends on the number of new firms

$$N_e = \frac{A_e}{N_e^{\psi}} Y_e^{1-\lambda} L_e^{\lambda}. \tag{A11}$$

The terms L_e, Y_e and A_e are the same as the baseline model but the new term N_e^{ψ} with $\psi > 0$ allows for entry costs to rise with the number of entrants in the equilibrium N_e. It captures congestion effects in Gutiérrez, Jones and Philippon (2021), Boar and Midrigan (2022a,b) and Walsh (2023). A positive ψ means that the resources needed per entry rise with the number of entrants in the equilibrium.

Real wage in this economy is given by

$$\ln w = \text{constant} + \frac{(\sigma - 1)\ln A + \frac{1}{1+\psi}\ln L + \frac{1}{1+\psi}\ln A_e}{\sigma - 1 - \frac{1-\lambda}{1+\psi}} \tag{A12}$$

Thus, the impact of A on variety and welfare is dampened when entry costs rise with productivity, either through higher labor costs (λ close to 1) or congestion (positive ψ).

B.3 A growth model with expanding varieties within firms

Consider our baseline growth model with only one region. Now we extend the model to allow for each firm to produce multiple varieties. In addition to choosing its quality A_t, each entering firm can also choose 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 and producing more varieties in a firm raises the entry cost of setting up the firm through $f(v_t, A_t)$:

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

As in the main text Profit maximization and free entry imply that

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

In addition, each firm's choice of the number for varieties satisfy

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

Since variable profits are $\pi_t(A_t, v_t) = \pi_t A_t^{\sigma-1} v_t$, 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

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

Suppose

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

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.[20] 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 real wage in this economy is

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

and the growth rate of the real wage is

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

Similar to the static love-of-variety model, a higher λ implies a smaller welfare effect of changes in the level and growth rate of A_t and L_t.

The equilibrium number of workers per firms is

$$\ln \frac{L_t}{N_t} = (\lambda - 1) \ln \frac{Y_t}{L_t} + \ln f(v_t, A_t) + \text{constant}$$

where N_t is the number of firms, where $\ln f(v_t, A_t)$ is constant due to the endogenous choice of A and v. The number of varieties produced in the economy is $M_t := N_t v_t$.

This model illustrates that amplification through entry of firms can occur in an endogenous growth model with rising quality, expanding variety, and

[20] We want to allow higher quality to facilitate growing variety per firm because there is evidence of variety growth in the U.S. See Bernard, Redding and Schott (2010) and Broda and Weinstein (2010).

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. Furthermore, the elasticity of the average worker per firm with respect to output per worker is informative of this amplification channel, even if firms per variety grows over time.

C Data

Table A1: Data sources

Data name	Source	Variables obtained
Business Dynamics Statistics (BDS)	U.S. Census	Firms (N), Establishments (N), Employment (L)
Gross State Product (GSP), Gross Domestic Product (GDP)	BEA / Haver	Y In 2012 constant \$
County Business Patterns (CBP)	U.S. Census	Alternative L
Commodity Flows Survey (CFS)	U.S. Census	$b_{s,s'}$ bilateral trade shares
Population	U.S. Census	Pop_t for estimation

Table A1 displays the data sources we use to construct the variables in the regressions. Our baseline regressions use the Business Dynamic Statistics (BDS) data from the Census Bureau, which is available yearly from 1978 to

2020 and the annual Gross State Product data (GSP) from the Bureau of Economic Analysis (BEA), which is available from 1977 onwards. The real GSP data from the BEA has a break in 1997 where the pre-1997 data is constructed using SIC industry level data in constant 1997 dollars while the post 1997 data is constructed using NAICS industry data in constant 2012 dollars. We use real GSP in 2012 dollars constructed by Haver Analytics from raw BEA data.

For the robustness checks, we additionally use the County Business Pattern (CBP) available yearly from 1986 to 2020. For the GMM estimations, we further add data on U.S. civilian non-institutionalized population and bilateral trade data from the Commodity Flows Survey (CFS). Both data are produced by Census Bureau.

The 1 year horizon regressions uses the change in employment per firm and output per worker between year $t-1$ and t and the change in the number of firms between year $t-2$ and $t-1$. The 41 year horizon specifications regress the change in employment per firm between 1979 and 2020 on the change in output per worker between 1979 and 2020 and the change in the number of firms between year 1978 and 2019. Similarly the 34 year horizon uses changes between 1986 and 2020 for output per worker and employment per firm and changes between 1985 and 2019 for the number of firms. For regressions with 10 year horizon, we calcualte changes over non-overlapping periods 1985–1996, 1997–2008, and 2009–2020.

D Additional empirical results

Table A2: Emp per estab. on GDP per worker and lagged number of estab.

National sample, 1978–2020

Dep variable	All plants	All plants	New plants	New plants
λ^{OLS}	1.203	0.982	1.108	-0.029
	(0.025)	(0.083)	(0.076)	(0.192)
ϕ^{OLS}		-0.237		-1.209
		(0.086)		(0.199)
R^2	0.612	0.663	0.047	0.501
N	43	42	43	42
Amplification	-6.3%	0.5%	-3.5%	18.4%
	(0.7%)	(2.2%)	(2.4%)	(2.3%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

Table A3: Average estab. size on GSP per worker and lagged number of estabs

Changes over time, U.S. states, 1978–2020

Horizon	41 years	41 years	1 years	1 year
λ^{OLS}	0.959	0.958	0.688	0.707
	(0.062)	(0.088)	(0.014)	(0.014)
ϕ^{OLS}		-0.005		-0.047
		(0.059)		(0.023)
R^2	0.004	0.005	0.194	0.173
N	100	50	2100	2050
Amplification	1.4%	1.4%	11.6%	10.3%
	(2.1%)	(3.0%)	(0.6%)	(0.6%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

Table A4: *New* estab. size on GSP per worker and lagged number of estabs

Changes over time, U.S. states, 1978–2020

Horizon	41 years	41 years	years	1 year
λ^{OLS}	1.203	1.208	0.679	0.670
	(0.105)	(0.127)	(0.106)	(0.111)
ϕ^{OLS}	0	0.260	0	-0.238
	0	(0.085)	0	(0.177)
R^2	0.037	0.222	0.004	0.006
N	100	50	2100	2050
Amplification	-6.4%	-8.6 %	12.0%	9.8%
	(3.1%)	(4.8%)	(4.4%)	(4.0%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

Table A5: Average emp per firm on GSP per worker and lagged # of firms

Changes over time, U.S. states, 1978–2020, for firms with 1 to 4 employees

Horizon	41 years	41 years	1 year	1 year
λ^{OLS}	1.081	1.048	0.980	0.955
	(0.029)	(0.026)	(0.010)	(0.010)
ϕ^{OLS}		0.134		0.165
		(0.016)		(0.014)
R^2	0.074	0.615	0.002	0.069
N	100	50	2100	2050
Amplification	-2.6%	-1.8%	0.6%	1.8%
	(0.9%)	(1.0%)	(0.3%)	(0.4%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.

Table A6: Emp per firm on GSP per worker and lagged number of Firms

U.S. states, 1986–2020, GSP per CBP worker instrument

Horizon	34 years	34 years	1 year	1 year
λ^{OLS}	0.955	0.950	0.788	0.791
	(0.102)	(0.098)	(0.015)	(0.015)
ϕ^{OLS}		0.136		-0.043
		(0.069)		(0.022)
N	50	50	1700	1700
Amplification	1.5%	2.0%	7.6%	7.2%
	(3.5%)	(4.0%)	(0.6%)	(0.6%)

Note: Employment and firms are from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau and the County Business Patterns. Real output is from the U.S. Bureau of Economic Analysis (BEA). λ^{OLS} is equal to one plus the regression coefficient on log output per worker and ϕ^{OLS} is equal to -1 times the coefficient on log lagged number of firms. Amplification refers to the indirect effect of increases in productivity A through increased entry (variety), and is equal to $\left(\frac{1-\lambda}{1-\phi}\right) / \left(\sigma - 1 - \frac{1-\lambda}{1-\phi}\right)$. We evaluate it at $\sigma = 4$.