Entry and Employment Dynamics in the Presence of Market Power

William Gamber*

May 10, 2022

Abstract

In this paper, I show that the link between business-cycle fluctuations in business formation and employment depends crucially on market structure. To do this, I study a general equilibrium model of producer dynamics in which producers' markups rise with their size, so that, in response to a decline in entry, incumbents' market shares increase and they increase their markups and reduce employment. In the model, a shock that leads entry to fall leads to a significant contraction in employment. These effects are significantly larger than in a model without variable markups.

*Contact: will.gamber@frb.gov. I am very grateful to my advisors Simon Gilchrist, Ricardo Lagos, and Virgiliu Midrigan and committee members Corina Boar and Mark Gertler for their guidance and support throughout this project. I would also like to thank Jaroslav Borovička, Ryan Decker, Giuseppe Fiori, Sebastian Graves, James Graham, Andrew McCallum, Dino Palazzo, Erick Sager, Michael Siemer, Venky Venkateswaran, and Joshua Weiss for their helpful comments, as well as seminar participants at NYU and the Federal Reserve Board. The views expressed in this paper are solely those of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System.

1 Introduction

During the Great Recession, the number of new businesses created each year declined by more than 35 percent relative to its peak in the mid 2000s and remained depressed through 2019.[1] This fall in entry accompanied a decline in employment of over 6 percent relative to trend that only slowly returned to its pre-recession level. An existing literature shows that the decline in entry can help explain the slow employment recovery.[2] In this paper, I show that link between business formation and aggregate employment depends crucially on the market structure of the economy. In particular, I show that fluctuations in entry matter more for aggregate employment when producers' market power increases more with market share.

My approach is to study fluctuations in entry in a general equilibrium model of producer dynamics in which market power varies with size. Producers in the model increase their markups as their market shares rise, so that a fall in entry leads incumbents to increase their markups and reduce employment. In the model, a fall in entry as large and persistent as the one experienced by the United States during the Great Recession leads the average markup to increase significantly and generates a decline in aggregate employment of 3 percent. I show that the same decline in entry leads employment to fall by 75% more in this model than in one in which producers' market power does not vary with their size.

This paper makes two primary contributions to the literature. First, I contribute to a literature that shows that business-cycle fluctuations in business formation generate persistent fluctuations in employment, both theoretically (Lee and Mukoyama (2015), Siemer (2014), Clementi and Palazzo (2016)) and empirically (Sedláček (2020)). I show that that the effects of fluctuations in entry depend on the market structure of the economy; when producers' market power increases with market share, a fall in entry has a larger effect on aggregate employment.

Second, I present and quantify a general equilibrium model with heterogeneous producers, entry and exit, adjustment frictions, and variable markups. The existing literature on entry over the business cycle assumes either that producers are homogeneous (Bilbiie, Ghironi and Melitz (2012) and Jaimovich and Floetotto (2008)) or that markups do not systematically vary with producer size (Clementi and Palazzo (2016), Lee and Mukoyama (2018), and Siemer (2014)). I show that incorporating variable markups implies larger and more immediate effects of entry on aggregate employment. I also show that in a model with producer heterogeneity, fluctuations in entry can have large effects on the aggregate markup in the presence of adjustment frictions. This

[1] Source: US Census Bureau Business Dynamics Statistics Database. Here, I define a business as an establishment, but similar statistics hold for firms.

[2] For recent examples, see: Sedláček (2020) and Clementi and Palazzo (2016).

result stands in contrast to a robust finding that entry fluctuations have small effects on the markup in frictionless models with producer heterogeneity (see, for example, Edmond, Midrigan and Xu (2018) and Arkolakis et al. (2019)).

I begin the paper by presenting motivating evidence for the key mechanism in my model, namely, that markups rise with producer size. My approach is motivated by the "production function approach" (PFA) that has been popular in the recent macroeconomic literature on markups (see De Loecker and Eeckhout (2017), for example). The intuition behind this approach is that, under the assumption that producers can frictionlessly adjust their variable inputs, the wedge between variable input use and revenue is informative about the size of the markup. I show that this wedge in the data varies strongly with producer size; the typical producer in the sample increases its variable input bill much less than one-for-one with its sales. This finding suggests that markups vary strongly with producer size.

I then present a general equilibrium Hopenhayn (1992) model with two key features: (1) a variable elasticity of demand and (2) labor adjustment costs. Producers in the model have ex-ante heterogeneous, stochastic productivity. They are each the monopolistic supplier of a differentiated variety and face downward sloping demand with an elasticity that declines with relative size. The shape of these demand curves implies that producers have an incentive to increase their markups as their output rises relative to the overall market. Producers must pay a convex hiring and firing cost, which slows their response to idiosyncratic shocks and prevents inputs from rapidly reallocating across businesses. Lastly, businesses exit each period and are replaced, in steady state, by newly created businesses.

I then use the quantified model to interpret the motivating evidence. As highlighted by Bond et al. (2020), the PFA requires the restrictive assumption that variable inputs can be costlessly adjusted. To relax this assumption, I use the reduced-form regression coefficients, along with data on employment reallocation, to discipline parameters in the model, including the degree of adjustment costs and the extent to which the elasticity of demand falls with producer size. I then simulate a panel of producers in the model and estimate the same PFA regressions on simulated data. I show that not accounting for adjustment costs leads to an overstatement of the relationship between producer size and markups but that large producers' markups do vary significantly with producer size, with an elasticity of around 25% to relative sales.

To study the effects of fluctuations in entry on aggregate employment, I then introduce a shock to the mass of potential entrants to the model. This shock can be interpreted as a shock to the cost to finance a new startup, and it leads to a reduction in entry. In the model, this temporary decline in entry has large and persistent effects on aggregate employment. The fall in entry increases the market shares of incumbent producers and leads them to increase their markups, produce less, and reduce employment. The most productive producers increase their markups the most, leading aggregate productivity to fall. These effects are economically significant; in response to a shock that reduces entry by one-third, as much as the fall during the Great Recession, the aggregate markup rises 0.75 percent and aggregate productivity falls 0.5 percent. Because of these changes, aggregate output falls 2.5 percent and employment declines 2 percent.

I next study the mechanisms in the model that generate these large fluctuations in employment in response to the fall in entry. I show that both adjustment costs and variable markups are key to generating this response, and a model missing either of these ingredients generates a much smaller increase in markups and decline in employment.

To study the role of variable markups in this model, I compare the model to one with a constant elasticity of demand. I find that the effects of entry on aggregate employment are 75 percent larger in the variable markups economy relative to the constant elasticity model. The difference between the two models arises because falling entry leads incumbent producers to increase their markups, leading to a decline in the labor share and a reallocation of output away from high-productivity producers in the variable elasticity model. I conclude that the existing literature on the role of entry in business cycle amplification understates the importance of business formation because it ignores the effects of entry on the markups of incumbents.

To study the role of adjustment costs, I next study a model with variable elasticity of demand but no adjustment costs. In that model, producers raise their markups in response to the shock to entry. This change in producer policy functions causes the unweighted average markup to rise. However, because small, low-markup producers face a higher elasticity of demand than large, high-markup producers, they benefit more from the fall in competition. This feature of demand implies that employment reallocates away from large producers to small producers, meaning that the employment-weighted average markup, the correct measure of the aggregate markup in this model, does not rise by much. Without adjustment costs, reallocation undoes 80 percent of the immediate rise in the markup. In the baseline model, adjustment costs prevent small producers from increasing their employment rapidly and inhibit this reallocation.

I conclude the paper by showing that secular trends in market power imply that entry matters much more for recessions now than it used to. I show that the within-producer correlation between variable input use and market share has fallen significantly since 1985; under the production approach assumptions, my estimates imply that the elasticity of the markup to revenue has more than doubled over the past 30 years. I account for this increase in the model with an increase in the rate at which the elasticity of demand changes with relative size. I show that this increase implies that entry fluctuations have larger effects on aggregate employment today than they used to.

2 Motivating facts

In this section, I motivate a key feature of the theoretical environment that I study: namely, that large producers increase their markups significantly as their market shares rise.

Data and sample

The data I use are a panel of publicly listed, US-based firms in Compustat. I restrict the sample to observations between 1985 and 2018, exclude financial firms and utilities, and for my baseline results classify firms using the Fama-French-49 industry definition.[3]

This sample, while not representative of the average firm in the economy, represents a large portion of US output and employment. Firms in this sample are only 1 percent of firms in the United States, but the sum of their sales is around 75 percent of nominal gross national income and their total employment accounts for 30 percent of nonfarm payrolls.

Markups and firm size

The measurement framework I use is motivated by the production function approach, recently popularized by De Loecker and Warzynski (2012). Consider a producer with a production function in a variable input L and a static input K.[4] The distinction between variable and static inputs is that the producer can costlessly adjust its variable input use, whereas its static inputs may be subject to adjustment costs. The producer's ability to create output and sell might depend on conditions out of the its control, such as productivity and demand, which I summarize with A. The production function can be expressed as

$$Y = Q(A; K, L). \tag{2.1}$$

Denote by α the output elasticity of the variable input L. This coefficient might

[3] This classification groups NAICS-4 industries by activity so that each group has roughly the same number of firms. The results that follow are not sensitive to the definition of industry – in Appendix A, I show that similar results hold using SIC and NAICS definitions at various levels of granularity.

[4] It is easy to extend this framework to the case with many variable and static inputs. In that case, the first-order condition that I derive below holds for *any* of the variable inputs.

vary over time or across firms and industries. A first-order condition with respect to L gives a relationship between total variable input cost WL, revenue PY, the markup μ, and the output elasticity.

$$WL = \alpha \frac{PY}{\mu}. \tag{2.2}$$

To estimate the relationship between the markup μ and revenue PY, I will then estimate how variable costs WL covary with revenue. Taking logs of this first order condition gives the following equation:

$$\log WL = \log \alpha + \log PY - \log \mu. \tag{2.3}$$

A larger covariance between markups and revenues at the firm level generates a lower value for β. If markups do not covary at all with revenues, then $\beta = 1$, and the more that this coefficient deviates from 1, the more that markups covary with revenue. To quantify how much firms increase their markups when their market shares rise, I estimate the following regression:

$$\Delta \log COGS_{ift} = \alpha_{it} + \underset{(0.002)}{0.654} \times \Delta \log PY_{ift} + \epsilon_{ift} \tag{2.4}$$

where ift denotes the observation for firm f in industry i at date t.[5,6] Consistent with the hypothesis that firms increase their markups as they grow, the estimated coefficient of 0.654 is statistically significantly less than one. Under the assumptions underlying the production function approach the estimated coefficient implies an elasticity of the markup to firm size of about 35%.

Relaxing the frictionless assumption

An alternative hypothesis for the less than one–for–one relationship between revenue and variable input use is the presence of variable input adjustment costs. These could be hiring and firing costs, long–term contracts in variable inputs markets, or other rigidities that inhibit a firm from increasing its variable input use when it faces a productivity shock. If a firm faced adjustment costs on its variable input (that is, it was not truly variable), then the static first order condition in the production function

[5] In parentheses, I show standard errors. See Appendix A.2 for more details on the regression, as well as estimates for a variety of specifications for the variable cost WL and choices of fixed effects.

[6] In contrast to De Loecker and Eeckhout (2017), I do not estimate the elasticity of output with respect to COGS, instead allowing fixed effects to pick up variation in α across firms and over time. This avoids two issues with their approach. First, I do not need to compute a measure of real output for each firm. As Bond et al. (2020) show, this is problematic in Compustat. And second, my approach allows for more heterogeneity in production functions across firms.

approach would not hold. In that case, the quantity μ represents any wedge distorting the firms' production choices away from their static optimums.

To understand how adjustment costs could lead to a less than one–for–one relationship between revenue and variable input use, consider a firm with an infinite labor adjustment cost. In response to an increase in productivity, the firm could increase its revenue without changing its employment at all, implying a regression coefficient of 0. The production function approach interpretation would mistakenly conclude that this firm increases its markups one-for-one with its relative size.

To avoid misattributing variation in this wedge to variation in the markup, I will use the model to discipline my interpretation of these regression coefficients. When I calibrate the model, I will jointly choose both the superelasticity of demand, which determines how market power varies with market share, and the degree of adjustment costs to match both the estimated coefficient in this regression and external data on firm–level labor adjustment dynamics. This strategy allows me to interpret these regressions in a structural model with adjustment costs. I will return to these regression estimates after I quantify the model.

3 Quantitative Model

In this section, I develop a general equilibrium producer dynamics model to study business cycle fluctuations in entry. The framework is a general equilibrium Hopenhayn (1992) model with a convex employment adjustment cost and variable elasticity of demand.

Environment

Time in the model is discrete and continues forever. There are three types of agents in this economy: (1) a representative household who consumes a final good, supplies labor, and holds a portfolio of all producers in the economy; (2) a final goods producer who uses a continuum of intermediate inputs to produce the final good; and (3) a variable measure of intermediate goods producers.

Household

A representative household chooses a state-contingent path for consumption of the final good $\{C_t\}$ and labor supplied $\{L_t\}$ to maximize the discounted sum of future utility:

$$\sum_{t=0}^{\infty} \beta^t u(C_t, L_t) \tag{3.1}$$

The household receives wage W_t and profits Π_t from its ownership of a portfolio of all producers in the economy. I normalize the price of the final good to 1, and so the household period budget constraint is:

$$C_t \leqslant W_t L_t + \Pi_t. \tag{3.2}$$

The intratemporal first-order condition of an optimal solution to the household's problem implies a labor supply curve:

$$W_t = -\frac{u_{L,t}}{u_{C,t}}. \tag{3.3}$$

Final goods producer

A perfectly competitive representative producer produces the final consumption good using a continuum of measure N_t intermediate goods as inputs. Each differentiated intermediate variety is indexed by ω. The final goods producer takes as given the prices of the intermediate goods and minimizes the cost of producing output. Its production function takes the following form:

$$\int_0^{N_t} \Upsilon\left(\frac{y_t(\omega)}{Y_t}\right) d\omega = 1, \tag{3.4}$$

where $\Upsilon(q)$ is a function that satisfies three conditions: it is increasing ($\Upsilon'(q) > 0$) and concave ($\Upsilon''(q) < 0$), and $\Upsilon(1) = 1$. Given quantities of each intermediate variety $\{y_t(\omega)\}$, aggregate output Y_t is defined as the solution to Equation (3.4).

The optimal solution to the cost minimization of the final goods producer implies a demand curve for each intermediate good:

$$p_t(\omega) = \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right) D_t. \tag{3.5}$$

where the aggregate quantity D_t is the demand index, defined as

$$D_t \equiv \left(\int_0^{N_t} \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right) \frac{y_t(\omega)}{Y_t} d\omega\right)^{-1}. \tag{3.6}$$

For the main exercises in this paper, I use the Klenow and Willis (2016) specification of $\Upsilon(q)$:

$$\Upsilon(q) = 1 + (\sigma - 1) \exp\left(\frac{1}{\epsilon}\right) \epsilon^{\frac{\sigma}{\epsilon} - 1} \left[\Gamma\left(\frac{\sigma}{\epsilon}, \frac{1}{\epsilon}\right) - \Gamma\left(\frac{\sigma}{\epsilon}, \frac{q^{\epsilon/\sigma}}{\epsilon}\right)\right] \tag{3.7}$$

where $\sigma > 1$, $\epsilon \geqslant 0$ and $\Gamma(s, x)$ denotes the upper incomplete Gamma function:

$$\Gamma(s,x) = \int_x^\infty t^{s-1}\epsilon^{-t}dt. \tag{3.8}$$

This specification of Υ generates an elasticity of demand for each variety that is decreasing in its relative quantity y_t/Y_t so that large producers set higher markups than small producers.[7] Under the Klenow and Willis (2016) specification,

$$\Upsilon'(q) = \frac{\sigma-1}{\sigma}\exp\left(\frac{1-q^{\frac{\epsilon}{\sigma}}}{\epsilon}\right) \tag{3.9}$$

In this case, the elasticity of demand is $\sigma q^{-\frac{\epsilon}{\sigma}}$. The demand elasticity declines with the quantity chosen of the intermediate good, and the elasticity of the elasticity of demand to quantity produced (the "superelasticity of demand") is the ratio $-\epsilon/\sigma$.

Intermediate goods producers

At each date t, a mass N_t of intermediate goods producers each uses labor to produce a differentiated variety. Each producer is the monopolistic supplier of a differentiated variety ω, and they hire labor in a perfectly competitive labor market at wage W_t. Each produces their variety using a constant returns production function $F(L;z) = zL$ and sells it to the final goods producer, taking as given their demand schedule.

Each period, each producer observes its idiosyncratic productivity z and the state of the aggregate economy, Λ. It then hires workers, produces output, and sells its differentiated variety to the final goods producer. Producers face labor adjustment costs $\phi(L, L')$ as a function of last period's employment L and their current employment L'. After selling their output and paying adjustment costs, each producer draws an i.i.d. fixed cost $\phi_F \sim G_F$ to operate in the following period. If it chooses not to pay the random fixed cost, it exits. The value of exit is normalized to 0. Producers are also forced to exit at rate γ. They discount future streams of profits using the discount factor m.[8]

[7] Similar forces exist in models of oligopolistic competition with a finite number of producers, such as Atkeson and Burstein (2008). However, this specification accommodates a continuum of producers and is a tractable way to model variable markups in a dynamic model without concerns about the existence of multiple equilibria in a dynamic game.

[8] In the deterministic steady state, the producer discounts future steams of profit at rate β, regardless of the household's stochastic discount factor. Later in the paper, I study deterministic dynamics. For my baseline results, I assume that producers discount future streams of profits using the risk neutral discount factor β. This assumption is equivalent to assuming either (1) the economy is small and open so its interest rate is fixed or (2) all producers are owned by a measure zero, risk-neutral mutual fund that distributes profits to households. The reason that I choose a risk-neutral discount rate is that the preference specification I use counterfactually implies that interest rates rise in recessions. As emphasized in Winberry (2020), interest rates are pro–cyclical, consistent with a countercyclical discount factor. In this paper, as in Winberry (2020), the interest rate affects producer dynamics. To avoid mischaracterizing the effect of falling entry on aggregate employment, I fix the discount rate and thus the interest rate.

Let Λ summarize aggregate states that are relevant to each producer. The recursive problem of an incumbent establishment who employed L employees last period, and has drawn productivity z is :

$$V(L,z;\Lambda) = \max_{p,L'} \pi(z,L',p;\Lambda) - c(L',L) + \int \max\left\{0, \tilde{V}(L',z,c_F;\Lambda)\right\} dJ(c_F), \tag{3.10}$$

$$\tilde{V}(L,z,c_F;\Lambda) = -c_F + \beta(1-\gamma)\mathbb{E}\Big[m'V(L,z';\Lambda)|z\Big], \tag{3.11}$$

$$\pi(z,L',p;\Lambda) = \left(p - \frac{W}{L}\right) d(p;\Lambda), \tag{3.12}$$

$$y \leqslant zL. \tag{3.13}$$

Equation (3.10) shows that the value of a producer is its period profits $\pi(\cdot)$, less the adjustment costs it pays $c(\cdot)$ and plus its continuation value. Its continuation value is the integral over fixed cost draws c_F of the value of continuing to operate next period. Equation (3.11) describes the value of continuing to operate, which equals the expected value of operating next period, discounted using the household's stochastic discount factor and the exogenous producer destruction rate, less the fixed cost of operation. Equations (3.12) and (3.13) describe the production function and demand system that producers face.

Entrants

Each period, a mass M_t of potential entrants considers whether to begin producing. Each potential entrant draws an idiosyncratic signal of their future productivity $\phi \sim F$ and decides whether to enter. After paying the sunk cost, the entrant freely hires labor but cannot produce. Its productivity the following period is drawn from a distribution $H(z|\phi)$.

The value of an entrant who has drawn productivity signal ϕ and already paid the sunk entry cost is

$$V_E(\phi) = \int_z \max_L \beta(1-\phi)\mathbb{E}\Big[V(z,L)|\phi\Big] dH(z|\phi). \tag{3.14}$$

In appendix G, I study the response of the economy to aggregate shocks when producers price streams of profit using the household's stochastic discount factor. In response to the decline in entry, consumption initially falls and returns to its steady state. Under the household preferences that I use, this movement leads the discount factor to fall. The decline in the discount factor has two effects that amplify the response of the economy to entry shocks: (1) it decreases the value of entry further and thus deepens and prolongs the fall in entry and (2) it makes producers more hesitant to hire.

The optimal policy of the potential entrant is to enter if and only if $c_E \leqslant V_E(\phi)$. Under regularity conditions about $H(z|\phi)$, the value function $V_E(\phi)$ is monotonically increasing in ϕ, and so the policy of the entrant is to enter if and only if its signal exceeds a threshold $\hat{\phi}$.[9]

Aggregation

There are useful aggregation results for this economy.[10] Consider the aggregate production function, where Z_t denotes aggregate productivity:

$$Y_t = Z_t L_t. \tag{3.15}$$

Some algebra shows that aggregate productivity is the inverse quantity–weighted mean of producer–level inverse productivities:

$$Z_t = \left(\int\int \frac{q_t(z,L)}{z} d\Lambda_t(z,L) \right)^{-1}. \tag{3.16}$$

This quantity grows with the number of producers (love of variety) and with the extent to which output is produced primarily by high–productivity producers. The superelasticity of demand is one source of misallocation, since it implies that large, high productivity producers restrict their output.

The aggregate markup is implicitly defined as the inverse labor share:

$$\mathcal{M}_t = \frac{Y_t}{W_t L_t}. \tag{3.17}$$

A rise in the aggregate markup implies a fall in the share of revenue paid to labor. One can show that the aggregate markup is the cost–weighted average of producer–level markups:

$$\mathcal{M}_t = \int\int \mu_t(z,L) \frac{\ell_t(z,L)}{L_t} d\Lambda_t(z,L). \tag{3.18}$$

4 Steady state

In the steady state of the model, producers are heterogeneous along a number of dimensions. Each producer's idiosyncratic state variables are its productivity and employment. Producers have a lifecycle, beginning small and slowly hiring workers

[9] An alternative to the selection model of entry presented here is free entry. In that model, the mass of potential entrants is unlimited, and each entrant decides whether to enter without observing any signal about their future productivity. In appendix H, I discuss this model and its implications for my results.

[10] Note that solving the model still requires approximating the value function of the producers. See Appendix E.1 for details.

and becoming more productive. Moreover, producers face labor adjustment costs, and so producers' output and pricing decisions are history dependent. In addition, producers differ in the elasticity of demand they face and thus in the markups they set.

Markups and producer size

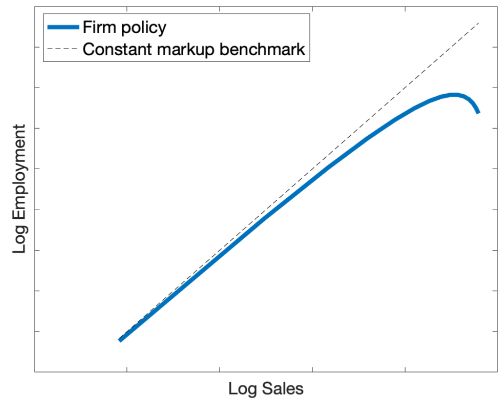
A key mechanism in the model is that the elasticity of demand falls with relative size, such that producers have an incentive to increase their markups as they grow relative to the market. To measure the strength of this mechanism, my calibration targets the regression coefficients I presented in section 2. Two mechanisms in the model directly affect this regression coefficient: (1) the elasticity of demand falls with producer size, leading producers to increase their markups as they grow, and (2) adjustment costs prevent producers from adjusting their variable input use in response to productivity shock.

To understand the role of mechanism (1), whose strength is dictated by the superelasticity, consider the model without adjustment costs. In that case the establishment's only idiosyncratic state variable is its productivity. As a its productivity rises, it produces more and its elasticity of demand falls. In response, it increases its markups. The increase in markups means that the producer increases its employment less than one-for-one with its sales. Figure 1 depicts the relationship between sales and employment in this model in blue, and the same relationship in a model with constant markups as the black dashed line. Producers in the variable markup model increase their markups as their sales grow, which implies that the slope of the sales-employment relationship is always less than one. Because larger producers increase their markups more with sales than small producers do, this relationship is also concave. For the largest producers, markups increase so much with sales that their employment actually falls as they gain market share.

While I estimate a linear regression of variable input growth on sales growth in the data, that relationship between employment and revenue is not linear in the model. This result presents a challenge in calibrating the model, as the average Compustat producer is larger than the average producer in the economy, which might lead me to overstate the extent to which markups rise with market share for the average producer. To calibrate the model, I choose parameters so that equation (2.4) estimated on a sample of the 1% largest producers in the model matches the regression from data. This procedure generates a comparable subsample to estimate the super-elasticity.

Adjustment costs also affect this regression coefficient: a higher adjustment cost leads input use to vary less with revenue. Because choosing two parameters to match one moment in the model presents an identification challenge, I also require the model

Figure 1: Employment and sales in the frictionless model



Note: The figure depicts the relationship between employment and sales in a version of the model with no adjustment costs. The constant markup benchmark (dashed black line) is a 45-degree line. Source: author's calculations

to match the autocorrelation of within-firm employment growth in Compustat. A higher value of the adjustment cost leads producers in the model to gradually respond to idiosyncratic shocks, increasing the autocorrelation.

Calibration

Functional forms

I use Greenwood, Hercowitz and Huffman (1988) preferences:

$$u(C_t, L_t) = \frac{1}{1-\gamma}\left(C_t - \psi\frac{L_t^{1+\nu}}{1+\nu}\right)^{1-\gamma}. \tag{4.1}$$

I also impose a quadratic form for the labor adjustment cost:

$$\phi(L, L_{-1}) = \phi_L\left(\frac{L-(1-\delta)L_{-1}}{(1-\delta)L_{-1}}\right)^2 L_{-1}.$$

I assume that productivity follows an AR(1) process in logs, with persistence ρ_z and innovation variance σ_z^2. The signal distribution for entrants follows a truncated Pareto distribution.

Calibration strategy

I fix six parameters and then jointly choose the remaining parameters to ensure that the model is consistent with salient features of the data. The pre-set parameter choices

Table 1: Pre-set parameters

Parameter	Description	Value	Source/Target
β	Discount factor	0.96	Annual model
$\mathbb{P}$(exit)	Probability of exit	0.11	Annual entry rate
M	Mass of entrants	1	Normalization
ν	Inverse Frisch elasticity	0.5	
δ	Job separation rate	0.19	

Note: This table summarizes part of the parameterization of the model. These parameter values were each chosen without targeting a particular moment in model simulations. Source: author's calculations.

are summarized in table 1. I then simultaneously choose productivity innovation persistence and dispersion ρ_z and σ_z, the adjustment cost parameter ϕ_L, the demand parameters σ and ϵ, and the Pareto parameter for the distribution of entrant signals ξ. To simplify the calibration procedure, I set the sunk cost of entry to 1 and the fixed cost of production to 0 with probability $(1 - \mathbb{P}(\text{exit}))$ and infinity with probability $\mathbb{P}(\text{exit})$.

While each of these parameters affects several moments in the model, each intuitively corresponds to one or two moments. The persistence of productivity and dispersion in its innovations affect the cross–sectional variance of producer–level log sales growth and the share of sales among the 10 percent largest producers. The Pareto coefficient affects the relative size of entering producers. I identify the degree of adjustment costs with the auto-correlation of producer-level log employment growth, which I estimate to be 12.81 percent in Compustat. A rise in the adjustment cost increases this auto-correlation; without the adjustment cost, the model generates a counterfactually negative auto-correlation. The superelasticity affects the relationship between producer size and the markup and so affects the within–producer regression coefficient of employment on sales. Table 2 summarizes the parameter choices.

The model performs well along a number of targeted and untargeted moments. Figure 3 summarizes the model's fit. As in the data, the model generates a wedge between labor and sales dynamism. The wedge between these two numbers is in line with that in the data. The model also fits the share of employment at entrant and young establishments that I estimate in the BDS. Fitting these variables is key to ensuring that the model accurately measures the aggregate importance of entrants. Finally, while the model matches the average cost–weighted markup of 1.25 that has been estimated in the data, it understates the value of the sales weighted markup, which is nearly 1.65 at the end of the sample in De Loecker and Eeckhout (2017). This disparity is likely due to the long right tail of sales in the data that is not present in a

Table 2: Calibrated parameters

Parameter	Description	Value	Targeted moment
ρ_s	TFP persistence	0.79	Top 10 percent share
σ_s	TFP innovation dispersion	0.18	Var. emp. growth
ϕ_L	Adjustment cost	0.07	Autocorr. emp. growth
ϵ/σ	Superelasticity	0.60	Labor–sales regression
ξ	Pareto shape of signal	0.95	Average size entering producer
σ	Elasticity parameter	20	Average markup

Note: Table summarizes part of the parameterization of the model. These parameter values were jointly chosen to match the 6 targeted moments. The variance and autocorrelation of employment growth and the regression coefficient were computed on a sample of the 1% largest producers in the simulated model economy. Source: author's calculations.

model with log-normal productivity.

Superelasticity estimate

My estimate of the superelasticity is consistent with estimates from a broad literature that uses producer–level data. Estimates of the superelasticity using microdata tend to be below 1. My estimates are close to Amiti, Itskhoki and Konings (2019), Berger and Vavra (2019), and Gopinath, Itskhoki and Rigobon (2010), who estimate the superelasticity using within-producer price responses to marginal cost shocks.

Consistent with other studies that use microdata to estimate the superelasticity, my value of $\epsilon/\sigma = 0.57$ is nearly two orders of magnitude smaller than estimates using macroeconomic data. As noted by Klenow and Willis (2016), the large estimates of the superelasticity needed to account for macroeconomic persistence are inconsistent with micro–level evidence. In this model, setting the superelasticity near the estimates in Lindé and Trabandt (2019) and Smets and Wouters (2007) would imply a counterfactually large markup-size relationship.

Market power versus labor adjustment

As discussed earlier, the within-producer regression coefficient of employment growth on sales growth could be less than 1 for several reasons. In the model, the two forces that generate the less-than-one-for-one regression coefficient are the positive superelasticity of demand and labor adjustment costs. The model allows me to decompose the reduced-form regression coefficient into each component.

The regression coefficient in the model is 0.628. When I set $\phi_L = 0$, re-solve the model, simulate a panel of producers in the new model, and estimate the regression

Table 3: Calibration targets and model fit

Moment	Target	Source	Model moment
$\text{Var}(\Delta \log L))$	6.17 percent	Compustat	5.8 percent
$\rho(\Delta \log L_t, \Delta \log L_{t-1})$	0.13	Compustat	0.1281
Labor–sales regression	0.654	Compustat	0.628
Average size of entering producer	50 percent	CP	0.52 percent
Frac. rel. sales. below 1	79 percent	Compustat	79 percent
Cost–weighted average markup	1.25	DLE	1.264
$\text{Var}(\Delta \log PY))$	14.15 percent	Compustat	13.4 percent
Top 10 percent share of sales	75 percent	Compustat	69 percent
$\rho(\Delta \log P_t Y_t, \Delta \log P_{t-1} Y_{t-1})$	0.12	Compustat	0.122
Share of employment at young producers	30 percent	BDS	32.97 percent

Note:DLEU: De Loecker et al (2019), CP: Clementi and Palazzo (2016)
Untargeted moments below line

Note: The table summarizes the model's fit of the data. It shows the targeted value and model moment. Explicitly targeted moments are above the single line. The variance and autocorrelation of employment and sales growth and the regression coefficient were computed on a sample of the 1% largest producers in the simulated model economy. Source: author's calculations.

coefficient, I find $\hat{\beta}_L = 0.65$. When I set the superelasticity of demand to 0, the regression coefficient rises to $\hat{\beta}_L = 0.92$. This decomposition suggests that labor adjustment costs account for between 9 percent and 20 percent of the deviation of the regression coefficient from 1.

This decomposition shows that ignoring variable input adjustment costs would lead an econometrician to overstate the relationship between firm size and market power. However, it also shows that, even accounting for variable adjustment costs, large firms' markups rise significantly with with their market shares; a conservative estimate attributes 8 percentage points of the elasticity of COGS to relative sales, implying an the elasticity of the markup to relative sales of 30%.

Aggregate parameters

There are two parameters whose values do not affect the steady state of the economy, only its response to aggregate shocks. These parameters are the inverse Frisch elasticity, which I set to be $\nu = 1/2$, and the disutility of labor parameter, ψ, which I set so that the steady state wage is 1.

Figure 2: Life cycle of the producer in the quantitative model

Note: The figure summarizes the lifecycle of an establishment in the model. Each panel shows the path of the average of a particular establishment-level variable for producers of a particular age relative to its average for all incumbents. Source: author's calculations.

The life cycle of the producer

Producers in the model, as in the data, start their lives small and grow slowly. Figure 2 shows that the average entering producer employs around 50 percent of the labor force of the average incumbent producer. They reach the size of the average producer by around age five. The model achieves this outcome in two ways: (1) the average productivity of entering producers is lower than that of incumbents and slowly reverts to the mean and (2) labor adjustment costs further slow the growth of new producers.

Producers' markups in the model also follow a life-cycle pattern, beginning low and slowly increasing.[11] The desire to set high markups derives from a demand elasticity that decreases with relative size. Because young producers' market shares slowly grow, their markups also increase slowly with age. The cost–weighted average markup increases by around 10 log points over the first five years of a producer's life in the model.

5 Shocks to entry over the business cycle

To study the role of variable markups in the transmission of fluctuations in entry to aggregate employment, I solve for the response of the model economy to a one-period unexpected shock to the mass of potential entrants. The size of the shock is chosen so

[11] Peters (2019) presents evidence for the lifecycle pattern of markups

that the resulting decline in the number of operating producers matches the decline in the number of establishments during the Great Recession.[12] After the initial shock is realized, all agents in the economy have perfect foresight of all aggregate variables going forwards as the economy returns to its steady state. I describe the solution method in more detail in appendix E.1.

I do not take a stance on the specific origin of the shock in the model, but it is consistent with hypotheses put forward in recent studies. The shock leads both the number of entrants and their average productivity to fall, consistent with a tightening of credit, as in Siemer (2019), or a fall in demand, as in Moreira (2016).

An entry shock

Figure 3 depicts the response of the baseline quantitative model to a shock to the cost of entry. The shock causes a fall in entry that leads the mass of establishments to decline by a little over 7 percent and the market shares of incumbents to rise. In response, incumbents increase their markups, and the cost–weighted average markup rises by 80 basis points. Because the labor share is the inverse of the average markup, it falls 80 basis points. Effective TFP, equal to the ratio of output to aggregate employment, falls gradually by nearly 1 percent. Employment falls 2 percent on impact, and output falls a bit over 2 percent. The wage satisfies the household labor supply equation and falls around 1 percent.

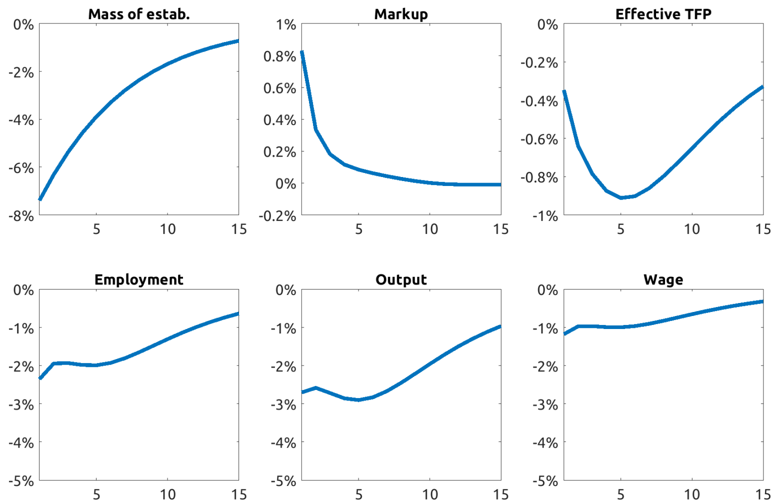
In response to the shock, the entry rate and share of employment among entrants and young producers fall. Figure 4 depicts the role of entrants following the shock. The entry rate falls by around 5 percentage points. It recovers quickly, with some overshooting, because the mass of entering producers recovers quickly while the mass of operating producers only gradually returns to its steady state level. The employment share among entering producers falls from 6 percent to around 3 percent.

Markups and productivity

To understand the roles of the average markup μ_t and aggregate TFP Z_t in generating the contraction in employment, it is useful to study the aggregated version of the model. This aggregated model is summarized by an aggregate production function (equation 5.1), the definition of the markup as the inverse labor share (equation 5.2), and the labor supply equation (equation 5.3).

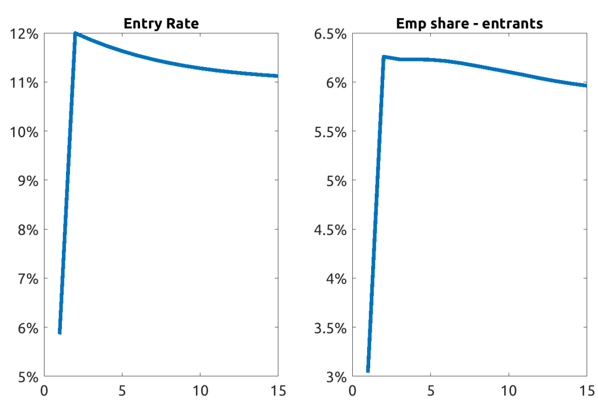
[12]While a shock to the cost of entry might seem more natural, in Appendix B I show that a shock to the cost of entry is inconsistent with the behavior of entrants during the Great Recession. In particular, a shock to the cost of entry leads the average size of entering producers to rise, which is not consistent with the Great Recession period.

Figure 3: Response of the baseline quantitative model to an MIT shock



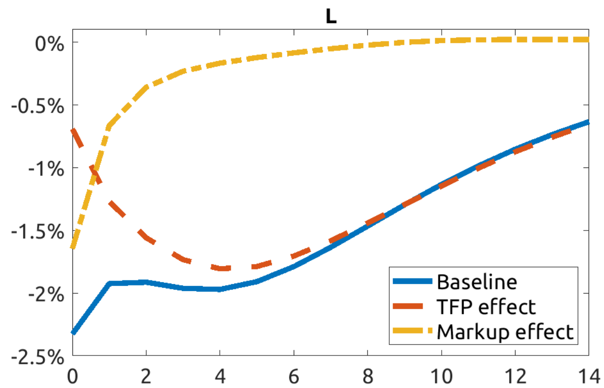
Note: The figure depicts the response of several aggregate variables to a one-time unexpected shock to the mass of potential entrants. Each line depicts the percent deviation of the variable from its steady state value. The size of the shock is chosen to match the fall in the number of establishments per capita during the Great Recession. The shock lasts for one period and the economy follows a perfect foresight path back to steady state. Source: author's calculations.

Figure 4: Entrants following the shock



Note: The figure depicts the path of the entry rate and employment share at entrants following the shock. Source: author's calculations.

Figure 5: Decomposition of entry shock



Note: The figure depicts a decomposition of the effects of the shock on aggregate employment into the effects of TFP and the effects of the markup. The dot-dashed line depicts the effect of the markup, holding aggregate TFP fixed. Each line depicts the percent deviation of the variable from its steady state value. The dashed line depicts the effects of TFP, holding the markup fixed. Source: author's calculations.

$$Y_t = Z_t L_t, \tag{5.1}$$

$$\mu_t = \frac{Y_t}{W_t L_t}, \tag{5.2}$$

$$W_t = \psi L_t^{\nu}. \tag{5.3}$$

Given paths for the cost–weighted markup μ_t and aggregate effective productivity A_t, equations (5.1) to (5.3) imply paths for output Y_t, employment L_t, and the wage W_t. While changing the paths of μ_t or A_t and recomputing these aggregate quantities does not necessarily represent an equilibrium of this economy, this representation of the economy allows for a decomposition of the response of aggregate variables to a shock.

Figure 5 depicts the paths of output, employment, and the wage under different paths for the markup and productivity. In blue, I allow both to follow their equilibrium paths. In red, I hold the markup fixed, and, in yellow, I hold TFP fixed. As they show, the rising markup generates a fall of 1.5 percent in employment, which represents most of the immediate decline in employment. As the markup gradually returns to its steady state value (with some overshooting), the decline in TFP accounts for all of the fall in employment.

The cost–weighted markup

The increase in the aggregate markup accounts for around one third of the contraction in employment. As discussed earlier, the relevant measure of the aggregate markup in this economy is the cost–weighted markup:

$$\mathcal{M}_t = \int\int \mu_t(z, L) \frac{\ell_t(z, L)}{L_t} d\Lambda_t(z, L) \tag{5.4}$$

The shock to entry affects the markups of individual producers $\mu_t(z)$ and the distribution of employment across producers. Two opposing forces affect the cost–weighted markup: (1) large producers raise their markups in response to the fall in entry and (2) there is a reallocation of employment from high–markup to low–markup producers.

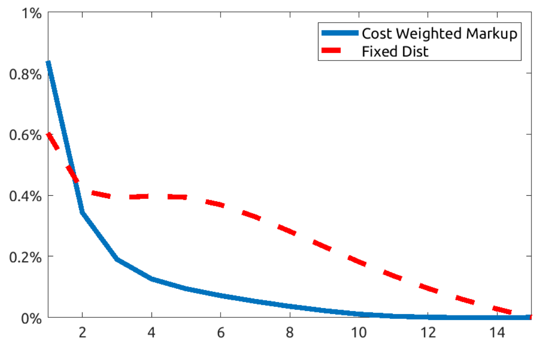
Adjustment costs slow the reallocation to low-markup producers. One way to see this phenomenon is to compare the path of the markup holding $\ell_t(z, L)/L \times d\Lambda_t(z, L)$ fixed. In the left panel of Figure 6a depicts this comparison. In red, I allow markups to vary but hold the distribution of employment fixed. This plot shows that the average producer raises its markups persistently in response to the shock. The black solid line shows the path of the markup in the baseline model and exhibits a more rapid return to its steady–state level. Following the shock, there is reallocation of employment to small, low-markup, producers.

The role of adjustment costs

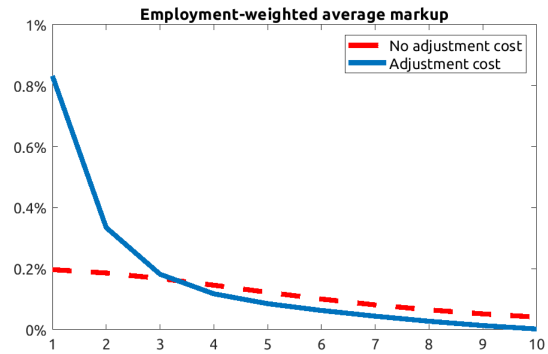
Adjustment costs slow the reallocation of output to low-markup producers. To quantify this mechanism, I compare the response in the baseline economy to the response in an economy without adjustment costs. In figure 6b, I plot the path of the cost-weighted average markup in each economy. As shown, without adjustment costs, the markup rises by 75 percent less than with adjustment costs.

Why does employment reallocate toward low-markup producers in response to the shock? The cause of variable markups in the model is a variable elasticity of demand; small producers set lower markups because they face a higher elasticity of demand than large producers. This feature also means that small producers are more exposed to competition from new entrants, and so they benefit more from the reduction in entry. Without adjustment costs, small producers' employment grows relative to that of large producers, leading the cost-weighted markup to increase only slightly. In this model, adjustment costs imply that small producers are not willing to hire rapidly, and so output is not reallocated as strongly to those producers.

Figure 6: Role of the adjustment cost in reallocation



(a) Weighted and unweighted markups



(b) Weighted markup and adjustment costs

Note: The left panel depicts the path of the cost-weighted average markup in response to the shock to the mass of potential entrants. The solid line depicts the path of the cost-weighted markup, allowing both the policy function of producers and the distribution of employment across producers to vary. The red dashed line shows the markup, holding fixed the distribution of employment across producers. The right panel depicts the path of the cost-weighted average markup in response to the shock to the mass of potential entrants in two different economies. The solid line depicts the path of the cost-weighted markup in the baseline economy. The red dashed line shows the same quantity in an economy without adjustment costs. Source: author's calculations.

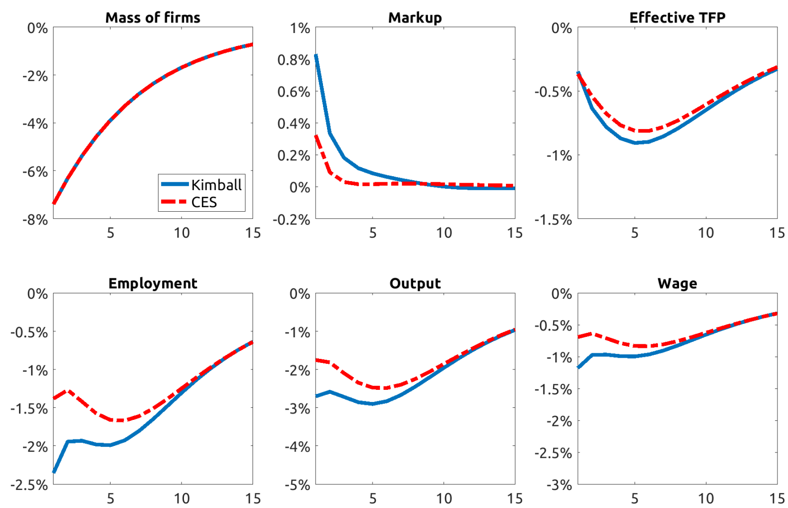
The role of variable markups

To quantify the role of variable markups in the propagation of entry shocks to aggregate employment and output, I compare the Kimball model to one in which producers' demand elasticities do not vary with their market shares. This comparison model features constant elasticity of substitution (CES) preferences. To ensure that the models are comparable, I choose the elasticity of substitution in the CES model so that the cost–weighted markup in each model is identical. I keep all other parameters the same.

I subject each economy to the same entry shock as before. Figure 7 depicts the results of this experiment. These impulse response functions show that variable elasticity of demand generates a significant fall in employment and amplifies the effects of an entry shock. The markup rises somewhat (by 37 basis points) in the CES model because adjustment costs push producers away from their frictionless optimal solution.[13] However, the rise in the markup is only about half of the rise in the Kimball model, meaning that employment in the CES model does not fall as sharply as in the Kimball

[13]Recall that the markup is price over marginal cost. The wage falls in response to the shock, giving producers an incentive to increase their employment. In a frictionless economy, firms would hire more workers and lower their price to keep their markups fixed. However, with an adjustment cost, not all producers will fully adjust in response to the shock, leading them produce less than they would absent the adjustment friction. Firms are constrained to be on their demand curves, and so this lower level of production is associated with higher prices.

Figure 7: Entry shock in the Kimball and CES models



Note: The figure depicts the path of the cost-weighted average markup in response to the shock to the mass of potential entrants in two different economies. Each line depicts the ratio of the variable to its steady state value. The solid line depicts the path of the cost-weighted markup in the baseline economy. The red dashed line shows the same quantity in an economy without adjustment costs. Source: author's calculations.

model. The additional rise in the markup in the Kimball economy generates a nearly 75 percent extra fall in employment on impact in the model with variable markups. This difference disappears after around five years.

6 The rising importance of markups for business cycles

A key moment in the quantitative model I study is the correlation between markups and producer size. In this section, I hypothesize that this relationship has grown stronger over time and show evidence for this hypothesis. I then show that entry shocks matter significantly more in a model consistent with the behavior of producers in data in 2015 than in 1990.

Variable input use and sales over time

One implication of a rising correlation between markups and producer size is that the estimated coefficient in equation (2.4) should fall. To study whether this is the case, I estimate this specification using a sequence of samples covering rolling centered five-

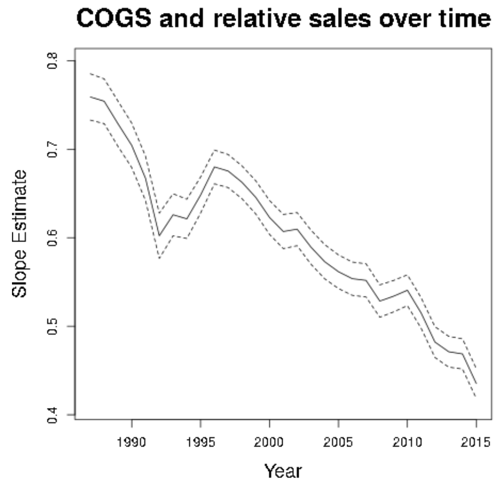


Figure 8: The within-firm relationship between COGS and relative sales over time

Note: The figure depicts the value of the estimated slope coefficient of a regression of within-firm log cost-of-goods-sold growth on log sales growth over time. Each point in the solid line represents the value of the regression estimated on a 5-year window centered around the date represented by that point. Dashed lines show 95% confidence bands. Source: Compustat and author's calculations.

year windows. Figure 8 summarizes the results this estimation. The coefficient declines significantly from 1985 to 2015.[14] Under the assumptions of the production function approach, the elasticity of the markup to relative sales rose from only 20 percent in 1990 to 55 percent in 2015.

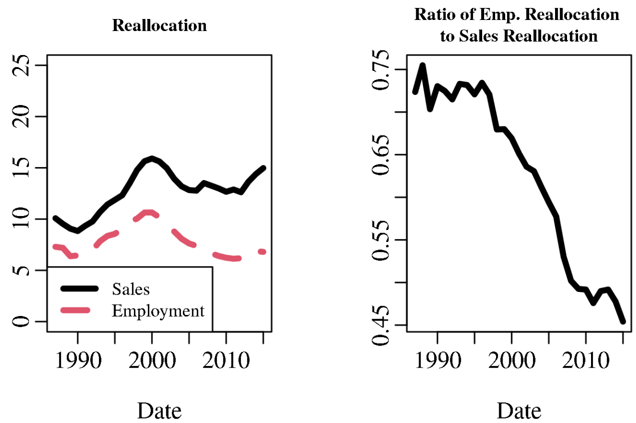
Markups and labor reallocation

Another testable implication of the hypothesis that the relationship between markups and market share has grown stronger is that the dispersion in firm-level employment growth has fallen relative to the dispersion in firm-level sales growth. To see why, I use the first-order condition from the production function approach discussed earlier to decompose the cross–sectional variance of sales growth ("sales reallocation") into the variance of employment growth ("employment reallocation") and two terms relating to markup variation:

$$\underbrace{\mathrm{Var}(\Delta\log PY)}_{\text{Sales reallocation}} = \underbrace{\mathrm{Var}(\Delta\log WL)}_{\text{Employment reallocation}} + \underbrace{\mathrm{Var}(\Delta\log\mu) + 2\mathrm{Cov}(\Delta\log\mu, \Delta\log L)}_{\text{Markup variation}}. \quad (6.1)$$

[14]See Appendix A.4.1 for a table summarizing alternative measures of variable costs and alternative specifications. All specifications show a decline in the relationship between producer size and variable input use.

Figure 9: Employment and Sales Reallocation



Note: The left panel depicts sales and emplyoment reallocation from 1985 to 2015. The right panel depicts the ratio of employment reallocation to sales reallocation over the same period. Source: Center for Research in Security Prices, CRSP/Compustat Merged Database, Wharton Research Data Services, http://www.whartonwrds.com/datasets/crsp/; author's calculations.

This decomposition shows that there is a relationship between the cross-sectional dispersion in labor and sales growth, mediated by markup dispersion. A positive markup-size relationship implies a wedge between sales and employment reallocation, and a stronger correlation between markups and producer size would lead to a decline in employment reallocation relative to sales reallocation.

I compute these measures in Compustat and find that labor reallocation has fallen relative to sales reallocation since 1990. The left panel of figure 9 shows that labor reallocation fell while sales reallocation remained roughly stable. The right panel shows the ratio of labor reallocation to sales reallocation over the same period. While employment reallocation used to be around 75 percent of sales reallocation, the ratio of the two has now fallen to 45 percent.

Implications for business cycles

I now study entry shocks in the model under two different calibrations, one that matches the 1985 regression values and the other that matches the 2015 values. I show that the secular change in the regression coefficient implies that aggregate employment responds more to fluctuations in entry than it used to.

Table 4: Selected moments, 1985 versus 2015 calibration

Calibration	ϵ/σ	β_L	Labor rea./Sales rea.	Cost-weighted markup
1985	0.455	0.77	60 percent	1.237
2015	0.7	0.468	26 percent	1.259

Note: The table describes several moments for two different calibrations. Each calibration corresponds to a particular value of the superelasticity, chosen to match the value of the employment-sales regression from that year. All other parameter values are equal to their baseline calibration values. Source: author's calculations,

Calibration

I choose the value of ϵ/σ to match the regression coefficient in 1985 of 0.786 and in 2015 of 0.486. As table 4 shows, the decline in this parameter generates a rise in the wedge between sales and labor reallocation, so that employment growth dispersion as a ratio of sales growth dispersion falls from 60 percent to 28 percent. This decline matches the decline of this ratio in data. So, the higher covariance between market share and markups implied by the regression coefficients can account for the rising wedge between sales and employment reallocation.

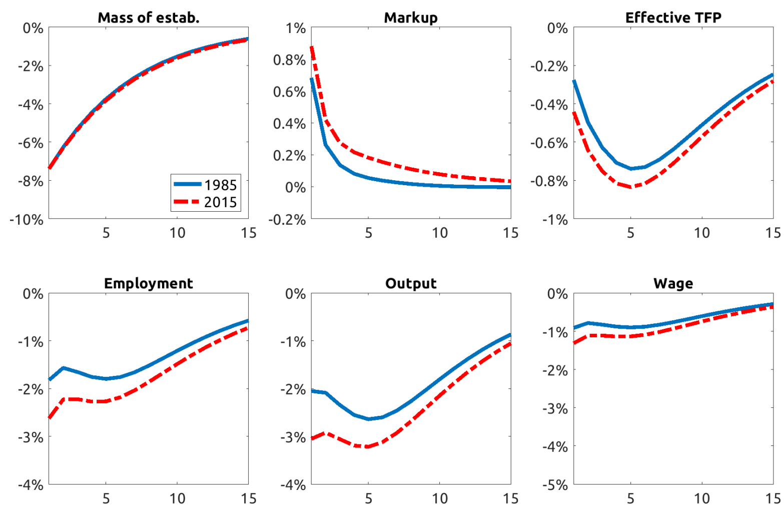
The rise in the superelasticity generates an increase in the cost-weighted markup of about 2 percentage points. This increase is about 20 percent of the actual rise in the cost–weighted markup, much of which, as De Loecker and Eeckhout (2017) notes, came from a reallocation of output to high-markup producers.

Impulse responses

Figure 10 depicts the response of each economy to the same transitory, unexpected shock to the mass of potential entrants. As it shows, the markup rises 75 basis points and only gradually recovers in the 2015 calibration, but in the 1985 calibration, it rises only 50 basis points and very quickly recovers. Effective TFP falls slightly more in the 2015 calibration. These two effects lead employment to fall 33 percent more in the 2015 calibration in response to the shock.

This exercise suggests that the rise in market power documented by De Loecker and Eeckhout (2017) and others might lead business cycles to become more volatile. As large producers' markups become more responsive to their market shares, fluctuations in entry will increase the volatility of aggregate employment.

Figure 10: Response to entry shock in 1985 and 2015



Note: The figure shows the response of six aggregate variables to the shock to the mass of potential entrants in two different calibrations of the model. Source: author's calculations.

7 Conclusion

Competitive conditions change dramatically in recessions. These changes were especially large during the Great Recession, when the number of operating establishments per capita fell by over 7 percent. Yet much of the recent literature on the effects of entry on the aggregate economy ignores the effects of entrants on the market power of incumbent producers. In this paper, I show that incorporating these effects into a general equilibrium, model of heterogeneous producers greatly amplifies the effects of entry on aggregate employment and output.

I first present a general equilibrium producer dynamics model with entry and exit, variable elasticity of demand, and adjustment frictions. I calibrate the model to be consistent with the life cycle of the producer, the adjustment costs of producers, and labor reallocation, as well as panel data estimates of a regression of variable input use on relative sales. I find that a fall in entry generates large falls in employment and output. The fall is nearly double relative to a model with constant markups.

I conclude with two quantitative applications of this model. In the first, I show that a sequence of shocks that generates the path of the number of establishments during the Great Recession in the model generates a persistent 5 percent decline in employment. In that simulation, employment returns to its steady state only by 2020. In the second application, I study the implications of the rise of market power for the effects of falling entry on markups. I show that the markup–size relationship in data has risen dramatically over the past 30 years. When I compare a model calibrated to the 1985 relationship to one calibrated to the 2015 relationship, I find that entry's effects on employment have increased substantially. This experiment suggests that rising market power amplifies the effects of entry on aggregate employment through the markup responses of large businesses.

There remain interesting avenues for future research. First, the countercyclical markups in the model may imply that inflation does not fall much in recessions. Future research could incorporate nominal rigidities into this model and study inflation dynamics. Second, what does optimal policy look like in this model? Is there a role for entry subsidies? How should the government treat large producers in recessions? Optimal policy is beyond the scope of this paper but is nonetheless relevant against the backdrop of the 2020 recession.

References

Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2019. "International Shocks, Variable Markups, and Domestic Prices." *Review of Economic Studies*, 86(6): 2356–2402.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *Review of Economic Studies*, 86(1): 46–80.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98(5): 1998–2031.

Berger, David, and Joseph Vavra. 2019. "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy*, 127(5): 2104–2142.

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

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2020. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data." National Bureau of Economic Research, Inc NBER Working Papers 27002.

Clementi, Gian Luca, and Berardino Palazzo. 2016. "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations." *American Economic Journal: Macroeconomics*, 8(3): 1–41.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–2471.

De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." National Bureau of Economic Research, Inc NBER Working Papers 23687.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2018. "How Costly Are Markups?" National Bureau of Economic Research, Inc NBER Working Papers 24800.

Gopinath, Gita, Oleg Itskhoki, and Roberto Rigobon. 2010. "Currency Choice and Exchange Rate Pass-Through." *American Economic Review*, 100(1): 304–336.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, 78(3): 402–417.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *The Review of Economics and Statistics*, 95(2): 347–361.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–1150.

Jaimovich, Nir, and Max Floetotto. 2008. "Firm dynamics, markup variations, and the business cycle." *Journal of Monetary Economics*, 55(7): 1238–1252.

Klenow, Peter J., and Jonathan L. Willis. 2016. "Real Rigidities and Nominal Price Changes." *Economica*, 83(331): 443–472.

Lee, Yoonsoo, and Toshihiko Mukoyama. 2015. "Entry and exit of manufacturing plants over the business cycle." *European Economic Review*, 77(C): 20–27.

Lee, Yoonsoo, and Toshihiko Mukoyama. 2018. "A model of entry, exit, and plant-level dynamics over the business cycle." *Journal of Economic Dynamics and Control*, 96: 1–25.

Lindé, Jesper, and Mathias Trabandt. 2019. "Resolving the Missing Deflation Puzzle." C.E.P.R. Discussion Papers CEPR Discussion Papers 13690.

Moreira, Sara. 2016. "Firm Dynamics, Persistent Effects of Entry Conditions, and

Business Cycles."

Peters, Michael. 2019. "Heterogeneous Markups, Growth and Endogenous Misallocation." *Econometrica*.

Sedláček, Petr. 2020. "Lost generations of firms and aggregate labor market dynamics." *Journal of Monetary Economics*, 111: 16–31.

Siemer, Michael. 2014. "Firm Entry and Employment Dynamics in the Great Recession." Board of Governors of the Federal Reserve System (U.S.) Finance and Economics Discussion Series 2014-56.

Siemer, Michael. 2019. "Employment Effects of Financial Constraints during the Great Recession." *The Review of Economics and Statistics*, 101(1): 16–29.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, 97(3): 586–606.

Winberry, Thomas. 2020. "Lumpy Investment, Business Cycles, and Stimulus Policy." Mimeo.

A Data

A.1 Cleaning procedure

I download a sample of Compsutat from WRDS. To clean the data, I use the following procedure:

- Keep only firms incorporated in the USA.
- Exclude utilities and financial firms – SIC codes 4900 - 4999 and 6900–6999.
- Exclude observations that are not in US dollars.
- Exclude observations with zero or negative values for SALE or EMP.

A.2 Other specifications

I estimate a variety of specifications for the variable cost WL and choices of fixed effects $g(ift)$. Table 5 summarizes the results. Each row contains results using a different measure of variable input cost, and in each column, I control for different levels of firm heterogeneity. I consider three measures of variable input use: total wage bill (XLR), total number of workers (EMP), and cost of goods sold (COGS). Data on wage bills are missing for many firms, and so I only have 17,501 observations of XLR, one-tenth the number of observations of COGS and EMP in the dataset.

Consistent with the hypothesis that firms increase their markups as their market shares grow, the estimated regression coefficient is statistically less than 1 across all nine specifications. My preferred specification is (3). In column (3), I estimate the regression using one-year growth rates.[15] This specification captures how, at a business cycle frequency, firms' variable input use varies when their revenues change relative to the whole industry. I find values well below 1 for these regressions, ranging from 0.356 for employment to 0.654 for COGS. These coefficients are interpretable as the amount by which a firm increases its variable input bill when its revenue growth is double that of the average firm in its industry.

[15] The results are robust to the definition of growth rate, but for my baseline results, I follow Haltiwanger, Jarmin and Miranda (2013) and use

$$g_{ift} = \frac{V_{if,t} - V_{if,t-1}}{\frac{1}{2}(V_{if,t} + V_{if,t-1})}.$$

Table 5: Variable input use and relative size over the whole sample

Dependent variable	(1)	(2)	(3)
$\log EMP$	0.8384	0.6275	0.356
	(0.0009)	(0.0016)	(0.0137)
$\log XLR$	0.8983	0.6716	0.4266
	(0.003)	(0.007)	(0.007)
$\log COGS$	0.9263	0.783	0.654
	(0.0007)	(0.002)	(0.002)
Specification	Log levels	Log levels	Growth rates
Fixed effects	Industry × Year	Firm + Industry × Year	Industry × Year

Note: This table depicts the results of estimating equation 2.4. Column (1) depicts the results using industry × year fixed effects. Column (2) depicts the results using firm + industry × year fixed effects. Column (3) depicts the results using growth rates. Source: Center for Research in Security Prices, CRSP/Compustat Merged Database, Wharton Research Data Services, http://www.whartonwrds.com/datasets/crsp/; author's calculations.

A.3 NAICS-4

In this section of the appendix, I document that the three facts are robust to using NAICS-4 as the definition of an industry.

Fact 1

Table 6: Variable input use and relative size over the whole sample

Dependent variable		log PY	
	(1)	(2)	(3)
$\log EMP$	0.8229186	0.623711	0.375305
	(0.0008742)	(0.001559)	(0.001798)
$\log XLR$	0.885107	0.688669	0.469273
	(0.003)	(0.005639)	(0.006349)
$\log COGS$	0.9164561	0.780266	0.651581
	(0.0007804)	(0.001595)	(0.001949)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

Fact 2

Table 7: Variable input use and relative size over time

Dependent variable	log PY		
	(1)	(2)	(3)
log EMP			
1986–1990	0.874916	0.565979	0.457095
	(0.002164)	(0.005299)	(0.004931)
2010–2014	0.802188	0.335218	0.261176
	(0.002643)	(0.005339)	(0.004834)
log XLR			
1986–1990	0.924773	0.70241	0.4436
	(0.004969)	(0.01274)	(0.0145)
2010–2014	0.821464	0.35053	0.29104
	(0.008911)	(0.02045)	(0.01651)
log $COGS$			
1986–1990	0.973087	0.793438	0.765169
	(0.001518)	(0.004944)	(0.004637)
2010–2014	0.911536	0.487565	0.504698
	(0.002448)	(0.007773)	(0.006566)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

A.4 NAICS-2

Fact 1

Table 8: Variable input use and relative size over the whole sample

Dependent variable		$\log PY$	
	(1)	(2)	(3)
$\log EMP$	0.8307641	0.632097	0.38278
	(0.0008417)	(0.001508)	(0.00174)
$\log XLR$	0.891063	0.683225	0.459426
	(0.002387)	(0.005004)	(0.005529)
$\log COGS$	0.9334514	0.79041	0.661271
	(0.0007165)	(0.00151)	(0.001869)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

Fact 2

Table 9: Variable input use and relative size over time

Dependent variable		$\log PY$	
	(1)	(2)	(3)
$\log EMP$			
1986–1990	0.873027	0.564924	0.449249
	(0.002279)	(0.005472)	(0.005122)
2010–2014	0.789511	0.329073	0.256887
	(0.002709)	(0.005524)	(0.004993)
$\log XLR$			
1986–1990	0.899926	0.71163	0.41474
	(0.006224)	(0.01455)	(0.01695)
2010–2014	0.80441	0.37426	0.30641
	(0.01006)	(0.02125)	(0.01752)
$\log COGS$			
1986–1990	0.956856	0.789263	0.760639
	(0.001668)	(0.005192)	(0.004856)
2010–2014	0.889245	0.47234	0.48915
	(0.002683)	(0.00817)	(0.00683)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

A.4.1 Regression results over time

Table 10: Variable input use and relative size over time

	log PY		
Dependent variable	(1)	(2)	(3)
log EMP			
1986–1990	0.888	0.585	0.483
	(0.002)	(0.005)	(0.005)
2010–2014	0.802	0.312	0.250
	(0.002)	(0.0.005)	(0.005)
log XLR			
1986–1990	0.926	0.57166	0.468
	(0.005)	(0.015)	(0.016)
2010–2014	0.812	0.222	0.261
	(0.001)	(0.025)	(0.021)
log $COGS$			
1986–1990	0.970	0.810	0.786
	(0.001)	(0.005)	(0.004)
2010–2014	0.900	0.466	0.486
	(0.003)	(0.008)	(0.007)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

Note: This table depicts the results of estimating each specification of equation 2.4 in years 1986–1990 and 2010–2014. Source: Center for Research in Security Prices, CRSP/Compustat Merged Database, Wharton Research Data Services, http://www.whartonwrds.com/datasets/crsp/; author's calculations.

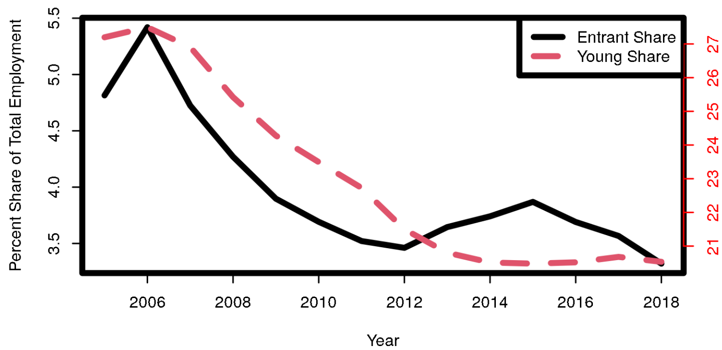
B Shocks to the cost of entry

In the main body of the paper I show the response of the economy to shocks to the mass of potential entrants. Another natural shock to study is to the cost of entry. As I show in this appendix, because of the selection mechanism present in this model, a shock to the cost of entry has very little effect on the employment share of entrants, which is inconsistent with the Great Recession period.

Figure 11 shows the employment share of entrants and young establishments during the Great Recession. Both fell sharply and persistently during the period. Figure 12 shows that following a shock to the cost of entry, the resulting decline in the entry rate has very little effect on the employment share at entering producers. The reason for the small imprint of declining entry on the employment share is that in the model, the marginal producer considering whether to enter is relatively unproductive and so employs few workers. A surprise increase in the cost of entry leads only producers marginal firms to decide not to enter and so has very little effect on employment at entering producers. By contrast, the shock to the mass of potential entrants leads the employment share at young establishments to fall in in line with the data.

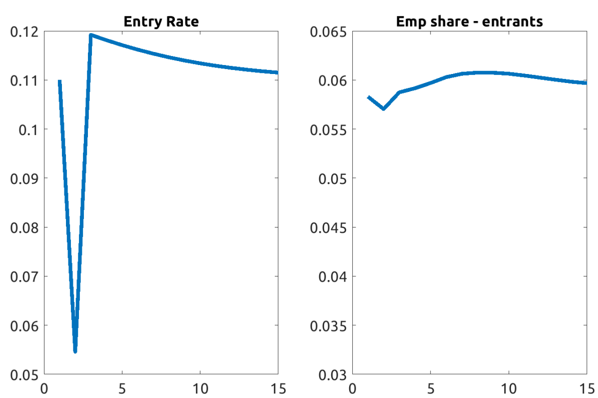
Because the entry cost shock only affects the smallest entrants in the economy, the shock has very little effect on the aggregate markup. Figure 13 shows the response of the model economy to the shock to the mass of potential entrants. The effects of the shock on the markup and effective TFP are greatly reduced, leading to a muted decline in employment and output.d

Figure 11: The Employment Share of Entrants and Young Establishments



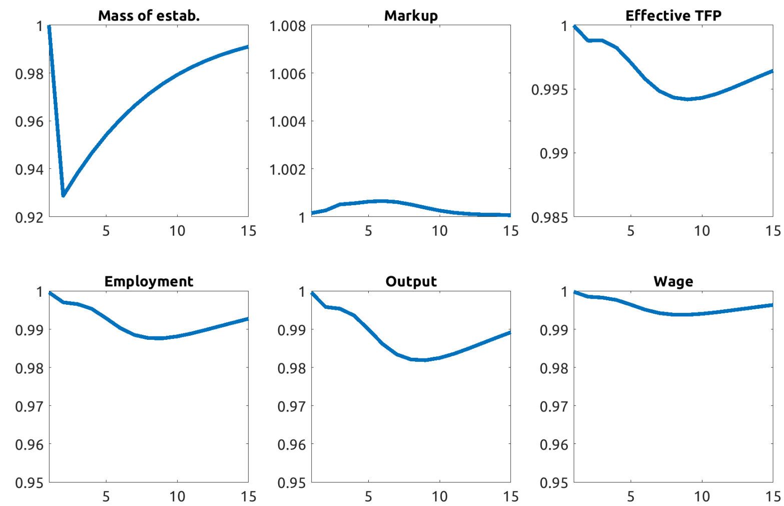
Note: The figure depicts the presence and behavior of entrants following a shock to the cost of entry in the model. The shock's size is chosen to match the fall in the number of operating establishments per capita during the Great Recession. Source: Census Bureau Business Dynamics Statistics database and author's calculations.

Figure 12: Behavior of entrants following a shock to the cost of entry



Note: The figure depicts the presence and behavior of entrants following a shock to the cost of entry in the model. The shock's size is chosen to match the fall in the number of operating establishments per capita during the Great Recession. Source: author's calculations.

Figure 13: Entry cost shock in the Kimball model

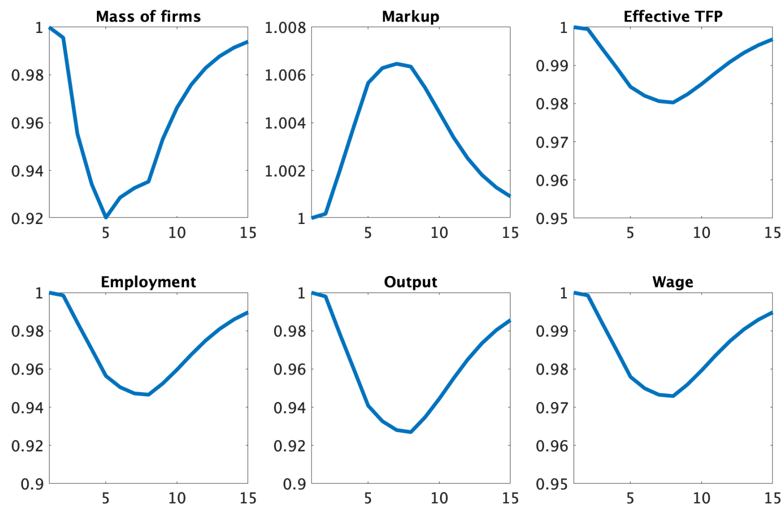


Note: The figure depicts the response of six aggregate variables to a shock to the cost of entry in the baseline model. Each line depicts the ratio of the variable to its steady state value. The size of the shock is chosen to generate an initial decline in the mass of establishments equal to the fall in the number of operating establishments per capita during the Great Recession. Source: author's calculations.

C Alternative calibration: firms

In this section, I study an alternative calibration in which the unit of analysis is the firm rather than the establishment. The key difference between the two calibrations is the average size of entrants. In the case of firms, entrants, on average, employ only 30% of the number of people as the average operating business. This reduces the effect of entry fluctuations. However, in the case of the Great Recession, the mass of operating firms fell by more relative to trend than did the mass of operating establishments. These second of these two effects dominates, and the effects of falling firm entry are slightly larger for firms than establishments during the Great Recession.

Figure 14: The Great Recession shock to the entry of firms



D Alternative calibration: Endogenous Exit

In this calibration, I allow for a non-degenerate distribution of fixed costs. This allows me to target the average size of exiting firms. As I show, this changes does not dramatically affect the results. Exit only varies slightly in response to shocks.

Table 11: Calibrated parameters

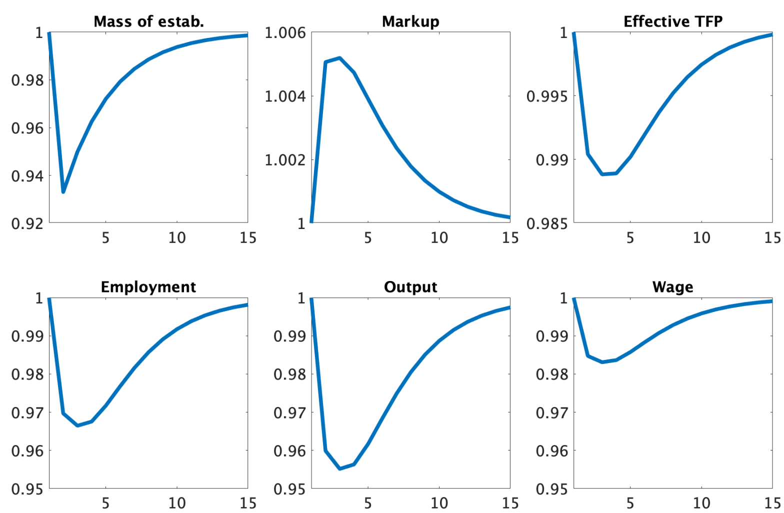
Parameter	Description	Value	Targeted Moment
σ_s	Tfp innovation dispersion	0.29	Labor Dynamism
ϕ_L	Adjustment cost	0.0032	Labor adjustment as fraction of revenue
ϵ/σ	Super-elasticity	0.6	Labor–sales regression
μ_F	Log fixed cost mean	-3.15	Entry rate
σ_F	Log fixed cost dispersion	1.65	Average size exiting firm
ξ	Signal Pareto tail	1.15	Average size entering firm
σ	Elasticity parameter	8.6	Average markup

Table 12: Calibration Targets & Model Fit

Moment	Target	Source	Model moment
Labor dynamism	7.5%	Compustat	4.97%
Sales dynamism	15%	Compustat	14.21%
Labor–sales regression	0.55	Compustat	0.57
Entry rate	11%	BDS	11.38%
Average size of exiting firm	59%	CP	58.92%
Average size of entering firm	50%	CP	49.39%
Cost–weighted average markup	1.25	DLE	1.255
Share of employment at entrants	6%	BDS	3.58%
Adjustment cost size	2.1 %	Bloom (2009)	1.81%
Share of employment at young firms	30%	BDS	37.03%

Note:DLEU: De Loecker et al (2019), CP: Clementi and Palazzo (2016)
Untargeted moments below line

Figure 15: The response of the baseline quantitative model to an MIT shock



E Solution method

E.1 Quantitative model

To find the initial steady state, I normalize aggregate output to 1 and the wage to 1. I approximate the value functions on a state space of a grid of 30 points for productivity and 50 points for labor. I discretize the productivity process using Rouwenhorst's method. Finding the steady state then involves finding a fixed point in the value of the demand index. The process is as follows:

1. Set D_L and D_U, the bounds on the values of the demand index.
2. Guess that $D_i = \frac{D_L + D_U}{2}$.
3. Given D_i, solve the value function of the incumbent firm. I solve this problem using value function iteration and the Howard Policy Improvement algorithm.
4. Given the value function of the incumbent firm, find the value of entry. This also implies policy functions of entering firms that depend on their productivity signal as well as entry decisions.
5. Given the policy functions of incumbent and entering firms, find the implied stationary distribution over the two state variables.
6. Compute the implied value of D_{out}. Define $diff = D_{out} - D_i$. If $|diff| < 10^{-8}$, the algorithm is complete. Otherwise, continue.
7. If $diff < 0$, then set $D_U = D_i$. Otherwise, set $D_L = D_i$. Return to step 2.

After completing this process, we can then fix a value that the Kimball aggregator should integrate to (note, for expositional purposes I use 1, but it is irrelevant as long as it is fixed) and a value ω such that the intratemporal first order condition of the representative household holds.

Solving for the response to an unexpected shock involves a shooting algorithm over W, C, and D.

F Pareto vs. Log-normal

.

Suppose, as in Edmond, Midrigan and Xu (2018), that firms face a static price-setting problem and that the distribution of productivity $G(z)$ is Pareto with minimum value 1. Denote by $q(z)$ and $\mu(q) = \frac{\sigma(q)}{\sigma(q)-1}$ the optimal policies of the firm. The cost–weighted markup in that case is

$$\mathcal{M} = \frac{\int_1^\infty \mu(q(z)) \frac{q(z)}{z} dG(z)}{\int_1^\infty \frac{q(z)}{z} dG(z)}$$

What do these optimal policies look like? The firm's optimal choice of q satisfies a first–order condition:

$$\Upsilon'(q) = \mu(q) \frac{1}{Az}$$

where A depends on the aggregate price index D and the price of labor, W. The more producers there are, the higher is W, and so an increase in entry (or an increase in N) increases W and decreases A. Also notice that the optimal choice depends on Az, not separately on A and z. We can then perform a change–of–variables $\tilde{z} \equiv Az$.

The Pareto assumption has convenient implications for the distribution $\tilde{G}(\tilde{z})$. To see why, assume z has location η and shape θ. Its CDF is then

$$G(z; \eta, \theta) = 1 - \left(\frac{\eta}{x}\right)^\theta$$

Performing the change of variables implies that:

$$\begin{aligned} G(\tilde{z}; \eta, \theta) &= 1 - \left(\frac{\eta}{Az}\right)^\theta && \text{(F.1)} \\ &= 1 - \left(\frac{\eta/A}{x}\right)^\theta && \text{(F.2)} \\ &= G(\tilde{z}; \eta/A, \theta) && \text{(F.3)} \end{aligned}$$

A change in A thus only affects the location of the Pareto distribution (up to rescaling). I show an example of this kind of shift in Figure 16

This implies that the markup then becomes:

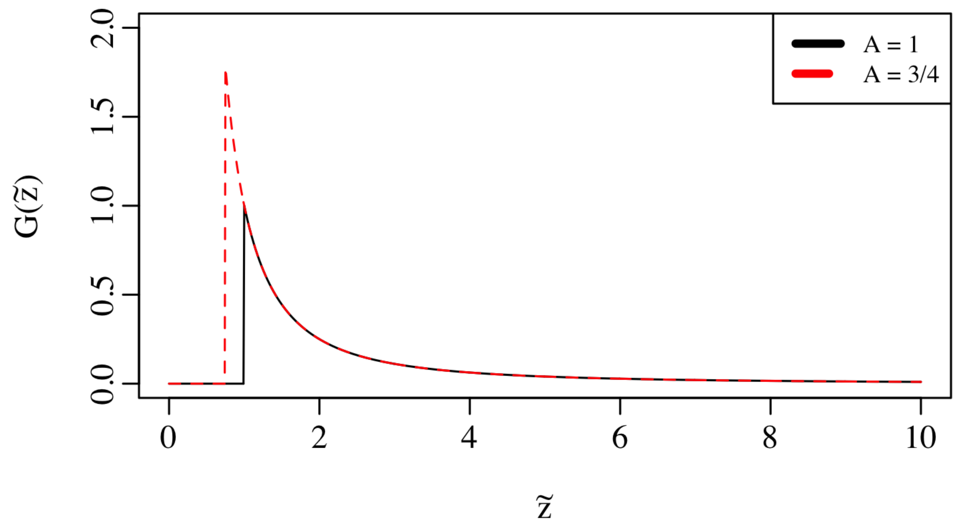
$$\mathcal{M} = \frac{\int_A^\infty \mu(q(\tilde{z})) \frac{q(\tilde{z})}{\tilde{z}} dG(\tilde{z})}{\int_A^\infty \frac{q(\tilde{z})}{\tilde{z}} dG(\tilde{z})}$$

Here I have used the fact that because z is Pareto distributed, so is $\tilde{z}$. A change in A only affects the lower bound of this integral. Since employment $\ell = q(z)/z$ is small at the lower bound of the integral, fluctuations in A only produce small fluctuations in $\mathcal{M}$.

What if instead we assume that productivity is log-normally distributed?

$$\mathcal{M} = \frac{\int_0^\infty \mu(q(z)) \frac{q(z)}{z} dG(z)}{\int_0^\infty \frac{q(z)}{z} dG(z)}$$

Figure 16: A change of variables under the Pareto assumption



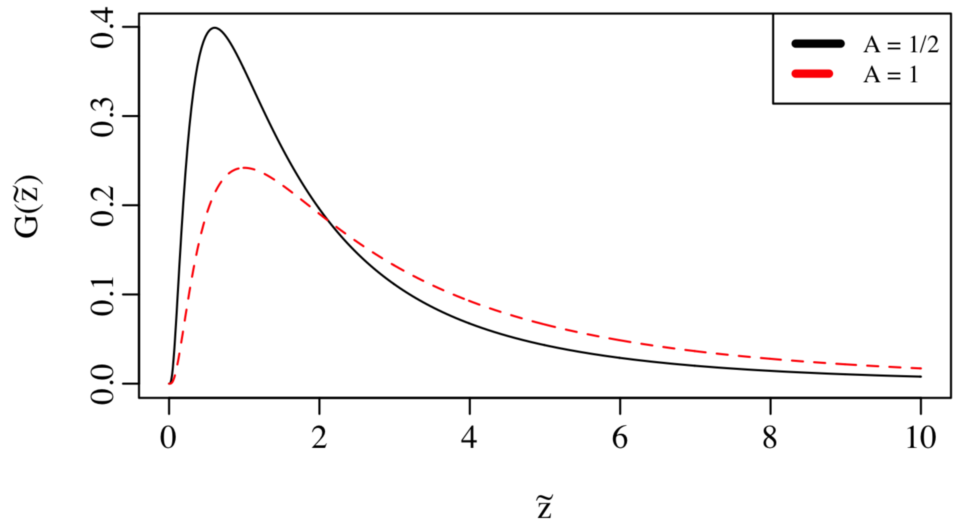
Suppose that $\log z \sim \mathcal{N}(\mu, \sigma^2)$. A change of variables implies that $\log \tilde{z} \equiv \log Az \sim \mathcal{N}(\log A + \mu, \sigma^2)$.

Recall the variance of a log-normally distributed variable:

$$\mathbb{E}[(\tilde{z} - \mathbb{E}(\tilde{z}))^2] = \exp(\sigma^2) - 1) \exp(2(\log A + \mu) + \sigma^2)$$

An increase in $\log A$ then increases both the mean and variance of $\tilde{z}$. Figure 17 depicts the effect of an increase in A on the distribution of effective productivity $\tilde{z}$. An increase in the variance of $\tilde{z}$ generally leads to a rise in concentration and an increase in the markup.

Figure 17: A change of variables under the log-normal assumption



G Stochastic Discount Factor

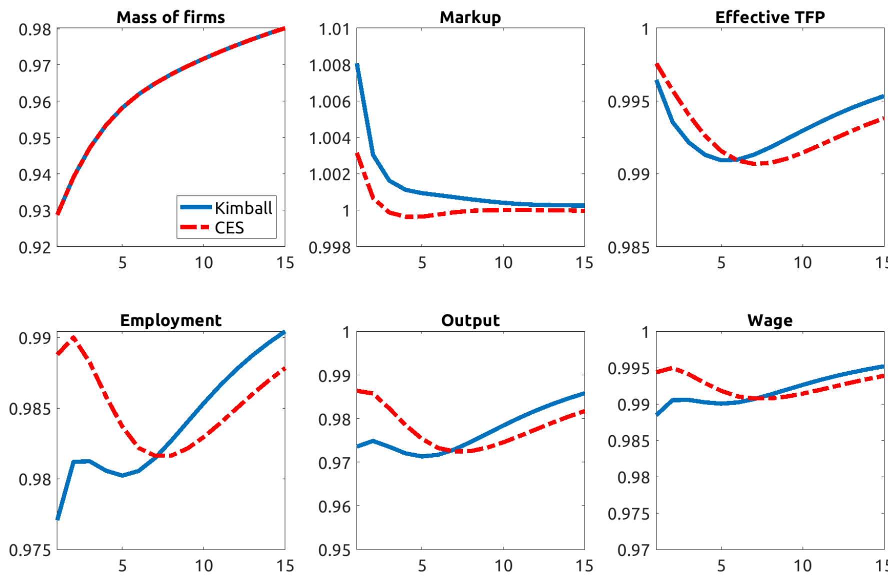
G.1 Shock to entry

In the case of Greenwood, Hercowitz and Huffman (1988) preferences, the stochastic discount factor is

$$m_{t+1} = \frac{\left(C_{t+1} - \psi \frac{L_{t+1}^{1+\nu}}{1+\nu}\right)^{-\gamma}}{\left(C_t - \psi \frac{L_t^{1+\nu}}{1+\nu}\right)^{-\gamma}}$$

I set $\gamma = 1$. The impulse response functions for the Kimball and CES economies to this shock are depicted in Figure 18. As they show, the variable SDF increases the persistence of the effects of the shock and the significance of the variable markups channel. The fall in the stochastic discount factor leads entry to fall by more. It also makes firms less willing to hire. These two effects lead to an increase in the persistence of (1) the decline in the mass of firms (2) the rise in the markup and (3) the fall in tfp coming from large firms producing less. These trends match the seemingly permanent nature of the shock to the mass of firms following the Great Recession.

Figure 18: Impulse response to an entry shock; variable stochastic discount factor



H Free Entry

An alternative to the selection model of entry that I use in the paper is free entry. With free entry, there is an unlimited mass of potential entrants each period. Each potential entrant decides whether to enter after observing the state of the aggregate economy and the entry cost but before observing any information about their idiosyncratic productivity. In equilibrium, these potential entrants will decide to become actual firms until the cost of entry exceeds the value of entry. Tables 13 and 14 summarize the calibration of the free entry model.

I solve for the response of the model economy to a one-time unexpected shock

Table 13: Calibrated parameters

Parameter	Description	Value	Targeted Moment
ρ_s	TFP persistence	0.79	Top 10% share
σ_s	Tfp innovation dispersion	0.17	Var. emp. growth
ϕ_L	Adjustment cost	0.055	Autocorr. emp. growth
ϵ/σ	Super-elasticity	0.57	Labor–sales regression
d_E	Productivity difference of entrants	0.4	Average size entering firm
σ	Elasticity parameter	20	Average markup

Table 14: Calibration Targets & Model Fit

Moment	Target	Source	Model moment
$\text{Var}(\Delta \log L))$	6.17%	Compustat	6.2%
$\text{Var}(\Delta \log PY))$	14.15%	Compustat	13.5%
$\rho(\Delta \log L_t, \Delta \log L_{t-1})$	0.13	Compustat	0.137
$\rho(\Delta \log P_t Y_t, \Delta \log P_{t-1} Y_{t-1})$	0.12	Compustat	0.116
Labor–sales regression	0.654	Compustat	0.0.656
Average size of entering firm	50%	CP	0.52%
Frac. rel. sales. below 1	79%	Compustat, industry average	79%
Cost–weighted average markup	1.25	DLE	1.25
Top 10% share of sales	75%	Compustat, industry average	68%
Share of employment at young firms	30%	BDS	32.9%

Note:DLEU: De Loecker et al (2019), CP: Clementi and Palazzo (2016)
Untargeted moments below line

to the cost of entry. The shock has persistence 0.685, the persistence of aggregate productivity in Clementi and Palazzo (2016). After the initial shock is realized, the all agents in the economy have perfect foresight of all aggregate variables going forwards as the economy returns to its steady state. I describe the solution method in more detail in Appendix E.1.

In response to the shock, the entry rate and share of employment among entrants and young firms fall. Figure 20 depicts the role of entrants following the shock. The entry rate falls by around 5 percentage points. It recovers quickly, with some overshooting, because the mass of entering firms recovers quickly while the mass of firms only gradually returns to its steady state level. The employment share among entering firms falls from 6% to around 3.5%.

Figure 21 depicts the paths of output, employment, and the wage under different paths for the markup and productivity. In blue, I allow both to follow their equilibrium paths. In red, I hold the markup fixed, and in yellow, I hold TFP fixed. As they show, the rising markup generates a fall of 1.5% in employment, most of the immediate decline in employment. As the markup gradually returns to its steady state value (with some overshooting), the decline in TFP accounts for all of the fall in employment.

Figure 19: The response of the baseline quantitative model to an MIT shock
Free Entry Model

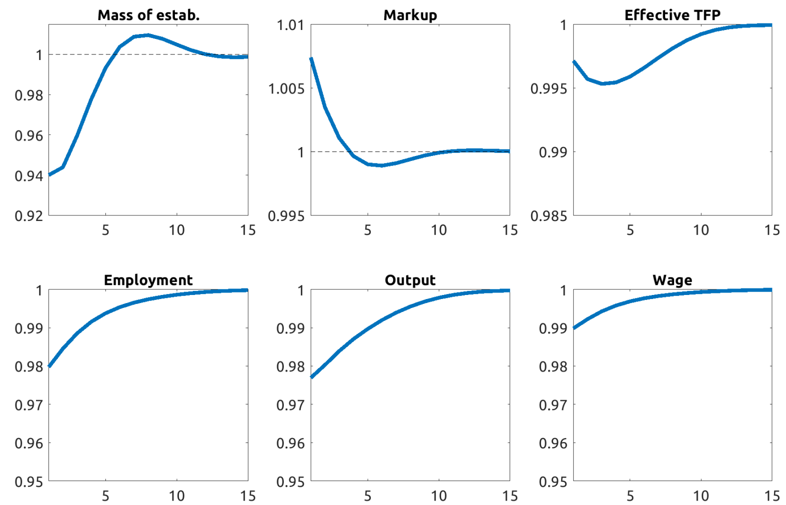


Figure 20: Entrants following the shock
Free Entry Model

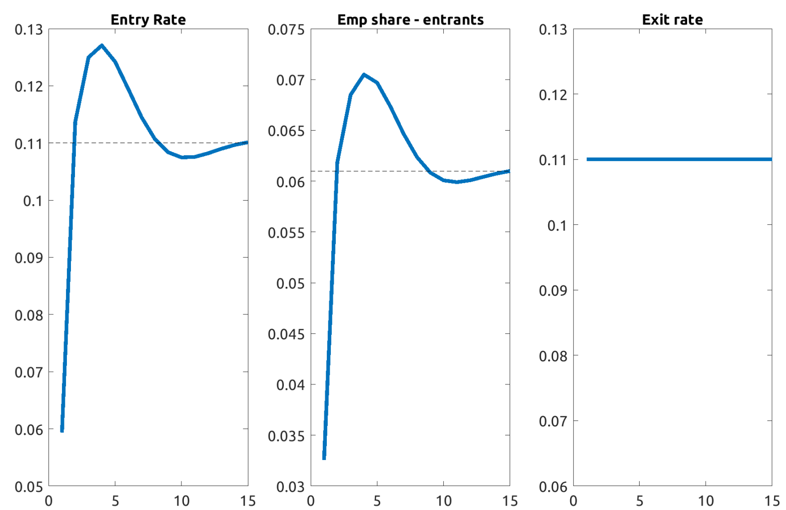


Figure 21: Decomposition of entry shock
Free Entry model