Entry, Variable Markups, and Business Cycles

Click here for most recent version

William Gamber*

September 15, 2021

Abstract

The creation of new businesses declines in recessions. In this paper, I study the effects of pro-cyclical business formation on aggregate employment in a general equilibrium firm dynamics model. The key features of the model are that the elasticity of demand faced by firms falls with their market share and that adjustment costs slow the reallocation of employment between firms. In response to a decline in entry, incumbent firms' market shares increase, their elasticity of demand falls, and they increase their markups and reduce employment. To quantify the model, I study the relationship between variable input use and revenue in panel data on large firms. Viewed through the lens of my model, my estimates imply that for large firms, the within-firm elasticity of the markup to relative sales is 25%. I use the calibrated model to study shocks to entry, finding that a fall in entry can lead to a significant contraction in employment. A shock to entry that replicates the decline in the number of businesses during the Great Recession generates a prolonged 2.5 percent fall in employment in the model. Finally, I show that the increasing correlation between market shares and markups over the last 30 years implies that the effect of entry on the business cycle has become stronger over time.

*Contact: will.gamber@frb.gov. I am very grateful to my advisors Simon Gilchrist, Ricardo Lagos, 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, 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% relative to its peak in the mid 2000s and remained depressed through 2018.[1] This fall in entry accompanied a decline in employment relative to trend of over 6 percent that only slowly returned to its pre-recession level. In this paper, I quantify the extent to which declines in the creation of new businesses amplify recessionary contractions in employment.

My approach is to study fluctuations in entry in a general equilibrium firm dynamics model. The model incorporates the idea that firms 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 as the one experienced by the US during the Great Recession leads the average markup to increase significantly and generates a decline in aggregate employment of 3 percent.

This paper makes three contributions. First, I present and quantify a general equilibrium model with heterogeneous firms, entry and exit, adjustment frictions, and variable markups. The existing literature on entry over the business cycle assumes either that firms are homogeneous (Bilbiie, Ghironi and Melitz (2012) and Jaimovich and Floetotto (2008)) or that markups do not systematically vary with firm size (Moreira (2017), 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. Second, I show that in a model with firm heterogeneity, fluctuations in entry can have large effects on the aggregate markup in the presence of adjustment frictions. This result stands in contrast to a robust finding that entry fluctuations have small effects on the markup in frictionless models with firm heterogeneity (See, for example, Edmond, Midrigan and Xu (2018) and Arkolakis et al. (2019)). My third contribution is empirical. I present a method for the quantification of the extent to which markups vary with firm size in the presence of adjustment frictions. This method innovates on the commonly-used production function approach, which requires the assumption of no adjustment costs (see De Loecker and Warzynski (2012) for a description of the approach, and Bond et al. (2020) for a discussion of its shortcomings).

I begin the paper by presenting 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 curves with an elasticity that declines with relative size. These demand

[1] The unit of analysis in this paper is the establishment, but similar statistics hold for firms.

curves imply that producers have an incentive to increase their markups as their output relative to the market increases. 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 next turn to microdata to quantify the key mechanisms in the model. I first present evidence that markups rise with firm size. My approach is motivated by the "production function approach" that has been popular in the recent macroeconomic literature on markups (see De Loecker and Eeckhout (2017), for example). The intuitive idea behind this approach is that, under the assumption that firms 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 firm size; the typical firm in the sample increases its variable input bill much less than one-for-one with its sales. Under the assumptions of the production function approach, my estimates imply that the typical firm in the sample increases its markup by 35 basis points for every 1 percent rise in its sales relative to the market.

I use the model to discipline the interpretation of these regression estimates. I choose parameters in the model, including the degree of adjustment costs and the extent to which the elasticity of demand falls with firm size, to ensure that it matches several moments in the data. I show that not accounting for adjustment costs leads to an overstatement of the relationship between firm size and markups but that large firms' markups do vary significantly with market share.

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 borrowing costs to finance new firms, 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 businesses and leads them to increase their markups, produce less, and reduce employment. The most productive firms increase their markups the most, leading aggregate productivity to fall. These effects are economically significant; in response to a shock that reduces entry by 1/3, as much as the fall during the Great Recession, the aggregate markup rises by 0.75% and aggregate productivity falls by 0.5%. Because of these changes, aggregate output falls by 2.5% and employment declines by 2%.

I next study the mechanisms in the model that generate these large fluctuations in employment in response to the fall in entry. My main finding is 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 50% larger in the variable markups economy relative to the constant elasticity model. The difference between the two models arises because falling entry leads incumbent firms to increase their markups, leading to a decline in the labor share and a reallocation of output away from high productivity firms in the variable elasticity model. I conclude that the existing literature on the role of entry in business cycle amplification understates the importance of firm entry 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, in response to the shock to entry, firms in the model raise their markups. This change in firm policy leads the unweighted average markup to rise. However, because small, low-markup firms face a higher elasticity of demand than large, high-markup firms, they benefit more from the fall in competition. This implies that employment reallocates away from large firms to small firms, 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% of the immediate rise in the markup. In the baseline model, adjustment costs prevent small firms from increasing their employment rapidly and inhibit this reallocation.

I conclude the paper with two applications of this theory. First, I study the persistent decline in business formation during the Great Recession. A shock to entry that replicates the decline in the number of establishments relative to trend over the period from 2007-2014 leads employment to decline by 3 percent, recovering to trend only in 2020. This exercise suggests that policies to extend credit to potential new businesses or to help cover the fixed costs of small businesses could have greatly accelerated the recovery out of the recession.

Second, in light of recent trends in market structure, I ask whether this channel has become more important over time. I show that the within-firm correlation between variable input use and market share has fallen significantly since 1985; my estimates imply that 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. It also implies that the standard deviation of employment growth has fallen relative to the standard deviation of sales growth, a fact that I confirm in the data.

Literature Review

The pro-competitive effects of entry

There is a long literature studying the role of entry in business cycle models. My approach is novel in that it incorporates both variable markups and labor adjustment costs into a general equilibrium business cycle framework that fully accounts for firm heterogeneity.

The idea that declines in entry during recessions might have anti-competitive effects is not new. An early literature studies this phenomenon in models in which firms are homogeneous (Jaimovich and Floetotto (2008) and Bilbiie, Ghironi and Melitz (2012)). It finds that fluctuations in entry have large effects on markups, productivity, and aggregate employment and output. However, heterogeneity is important and likely reduces the effects of entry on aggregates. Entering firms are significantly smaller on average than incumbent firms, which limits the effects of entry on the market shares of incumbents (Midrigan (2008)).

Even when entrants are the same size as incumbents, introducing heterogeneity into variable markups models reduces the pro-competitive effects of entry. A recent literature finds that entry has little to no effect on the aggregate markup in a class of economies with firm heterogeneity. This result is quite robust. Edmond, Midrigan and Xu (2018) study a model similar to mine, except with no adjustment frictions, and find that marginal changes in entry have approximately zero effect on the cost–weighted markup. This result holds for a simple reason. Small firms are most exposed to competition, and so while a fall in entry increases the markups of all firms, it also reallocates inputs away from high markup to low markup firms. In these models, the aggregate markup is the cost-weighted average of firm-level markups, and so the reallocation mechanism undoes the rise in the aggregate markup following a drop in entry. Similar results arise in Arkolakis et al. (2019) and Bernard et al. (2003).

While this reallocation may be relevant in the long run, it is inconsistent with the behavior of firms at business cycle frequencies. Inputs are not rapidly reallocated between firms during recessions, and there are frictions that prevent small firms from picking up slack labor demand form large firms. In fact, small firms' sales fall by more than large firms' in recessions (Crouzet and Mehrotra (2020)), and the share of employment at new and young firms fell sharply during the Great Recession. I modify the frictionless Pareto framework in two ways. First, I assume a log normal productivity distribution, and second, I include labor adjustment costs.[2] Labor adjustment costs prevent the extreme reallocation of employment to low markup firms from undoing the

[2] Arkolakis et al. (2019) find that the effects of entry on the markup do not differ significantly between the assumptions of Pareto and log-normal productivity.

firm-level increase in markups.

In this sense, my paper is an effort to quantitatively distinguish between the early literature's finding that entry has large pro-competitive effects in homogeneous firms models (Bilbiie, Ghironi and Melitz (2012), Jaimovich and Floetotto (2008)) and the neutrality results of the more recent literature (Edmond, Midrigan and Xu (2018) and Arkolakis et al. (2019)). My analysis takes firm heterogeneity into account, both with respect to size and age. I find that because of the limited role of reallocation across firms, there are sizable pro-competitive effects of entry at business cycle frequencies, and so, the relevant calibration of my model is closer to the homogeneous models of the early literature than the frictionless, heterogeneous firms models of the more recent literature.

My paper's findings are consistent with recent "reduced-form" causal evidence of the effects of entry on prices. Jaravel (2019) provides evidence that entry affects price setting behavior. He finds in grocery store scanner data that product categories with higher demand growth experience lower price growth. He rationalizes this finding by showing that higher demand growth product categories also experienced higher rates of new product creation. Felix and Maggi (2019) provides causally-identified evidence from a market reform in Portugal that increased entry leads aggregate employment to rise. Finally, in complementary work, Suveg (2020) studies the effects of exit on markups. Using an instrumental variables identification strategy, she shows in Swedish data that a one percent increase in exit generated by a reduction in the availability of financing led prices to increase by 1.6 percent.

My paper's finding that entry significantly affects aggregate economic activity is also consistent with Gutiérrez, Jones and Philippon (2019). They estimate a general equilibrium model of entry and exit using time-series and cross-sector variation in entry rates, output, investment, and Tobin's Q. They find that rising entry costs account for a 15 percentage point rise in the aggregate Herfindahl index and a 7 percent decline in the capital stock. Their model features constant markups and homogeneous firms and thus omits the key mechanism I study in this paper.

The Great Recession

A number of papers study the effects of entry on output and employment during the Great Recession. Siemer (2014) and Moreira (2017) both document that young firms start small and contribute significantly to aggregate employment growth. These papers argue that during recessions, there are forces (financial constraints in Siemer (2014) and demand constraints in Moreira (2017)) that limit the number and size of new firms. A lack of entry and the persistence of idiosyncratic conditions generate a "missing cohort" of firms, whose absence from the economy has long lasting effects.

Clementi and Palazzo (2016) study these effects in general equilibrium. In spite of the large variation in the economic presence of entering and young firms, they find that entry plays a surprisingly small role in propagating recessions. The key reason for this apparent contradiction is that, in general equilibrium, wages fall to induce incumbent firms to hire the workers who would have been employed at the missing entrants. This, coupled with the fact that entering establishments comprise only 5% of the economy's employment means that general equilibrium models of entry find only modest effects of the variation in entry on aggregate employment. In the model I study, large incumbent producers' elasticity of demand falls when entry falls, leading them to increase their markups and preventing them from picking up the slack in labor demand when the wage falls.

2 Entry over the business cycle

In this section, I use the Census Bureau's Business Dynamics Statistics database (BDS) to document empirical regularities about the role of entrants in the economy. I show that entry varies strongly over the business cycle and discuss the relative size of entering firms and establishments. The BDS is constructed from the Longitudinal Business Database, and it contains information about employment and the number of businesses at an annual frequency, aggregated by firm size and age. The dataset I use covers years 1977–2018.

Entry rates in the typical recession

The entry of new establishments falls in recessions and rise in booms, driving a pro-cyclical growth rate in the number of operating firms and establishments. Figure 1 shows the the annual log growth rate of the number of establishments each year in the BDS ("net entry"). Net entry is on average around 1 percent per year, but it fluctuates pro-cyclically. The 1980, 1981-1982, and 2007-2009 recessions exhibited particularly volatile fluctuations in the growth rate of the number of businesses, and the fall in the number of businesses during the Great Recession was especially large and persistent.

Pro-cyclical net entry is driven primarily by pro-cyclical gross entry rates. Figure 2 depicts firm entry and exit rates in the BDS. Average entry and exit rates have both declined substantially since 1980, though the change is more pronounced for entry. The right panel of Figure 2 depicts the data detrended using a 5-year trailing average. It shows that both entry and exit rates fluctuate relative to trend during recessions.

Given that these are aggregate fluctuations, they mask considerable heterogeneity in business dynamism across industries. They are, for example, muted relative to the fluctuations in manufacturing plants documented by Lee and Mukoyama (2015), who

Figure 1: Annual growth in the number of establishments per capita in the BDS

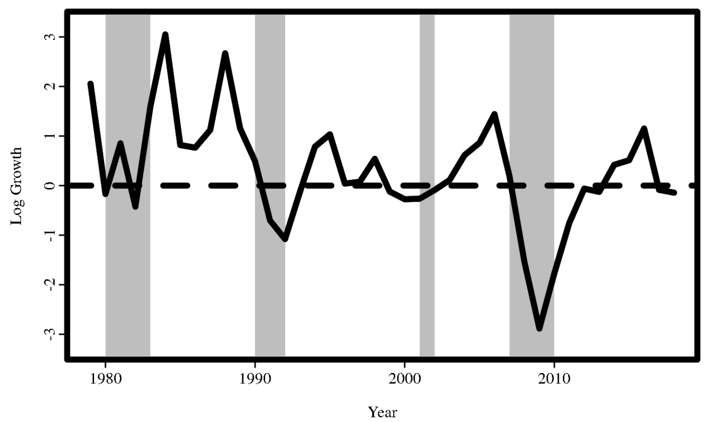


Figure 2: Entry and exit of establishments in BDS

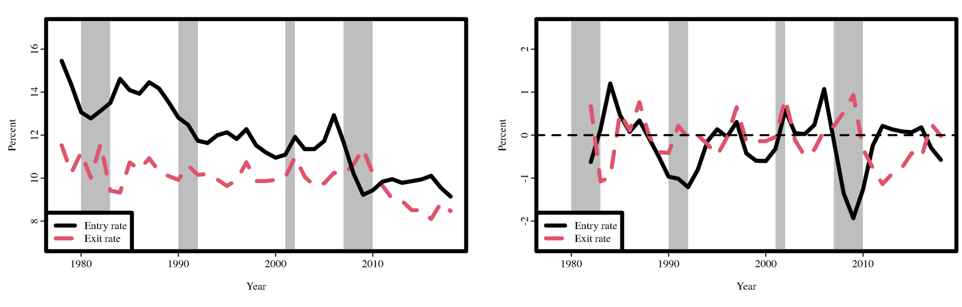


Table 1: Entrants relative to the whole economy, 1985–2006

Moment	Establishments
Entry rate	12.32%
Emp. share entrants	6.13%
Emp. share young	29.3%
Relative size of entrants	50.2%

find that entry rates are 4.7% lower in recessions than they are in booms. They also find that exit rates are only mildly procyclical, falling by 0.7% in recessions.

The employment share of entrants and young businesses

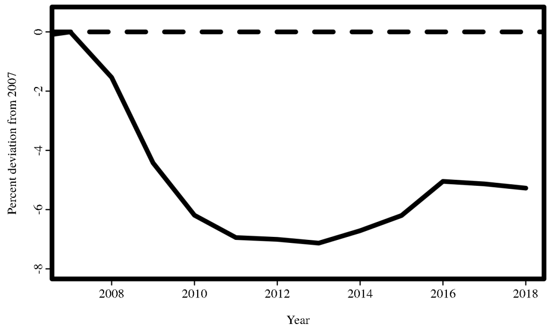
Entrants are smaller than incumbents on average. While entering establishments comprise roughly 10 % of total firms, they comprise only 6% of total employment, and the average entrant employs about half the number of people as the average establishment. These estimates from the BDS are consistent with the facts established in Lee and Mukoyama (2015) about manufacturing plants. They find that entering plants are 50% of the size of the average and exiting plants are around 35% of the size of the average. Table 1 shows similar facts in the BDS.

The Great Recession

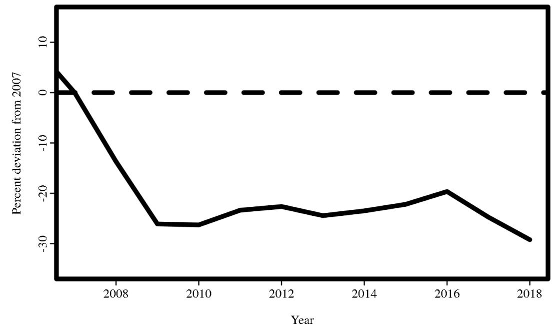
The fall in entry was particularly pronounced during the Great Recession. As panel (a) of Figure 3 shows, the number of operating establishments per capita fell gradually, reaching 7.13% below its 2007 peak in 2013 and only slowly recovering thereafter. Panel (b) shows a large and persistent fall in the number of operating establishments of between 20% and 30%, and panel (c) shows an increase in establishment exit through 2010 that then gradually declines through 2016. Panels (b) and (c) confirm that, while exit may have contributed to the short-run fall in the number of operating establishments, entry was the primary driver of the large and persistent fall in the number of operating establishments.

The employment shares of young and entering firms have been pro–cyclical since 1978, when the data begin, with the Great Recession exhibiting the largest and most persistent fall in the economic importance of young businesses. The share of employment at young establishments, for example, fell from around 30% in 2007 to nearly 20% by 2012. These large fluctuations in the presence of new businesses in the economy suggest a role for entry in business cycle propagation.

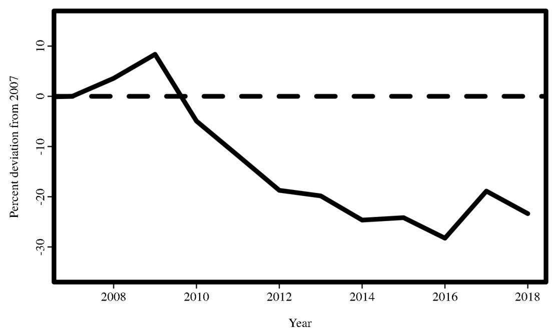
Figure 3: Establishments per capita during the Great Recession



(a) Total estab. per capita

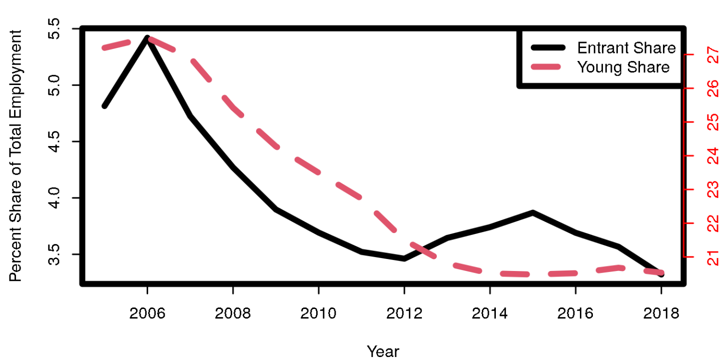


(b) Entrant estab. per capita



(c) Exiting estab. per capita

Figure 4: Employment share of young and entering businesses



3 Quantitative Model

In this section, I develop a general equilibrium firm 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 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 firms 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 firms 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 firm produces the final consumption good using a continuum of measure N_t of 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 ($\Upsilon'(q) > 0$), 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:

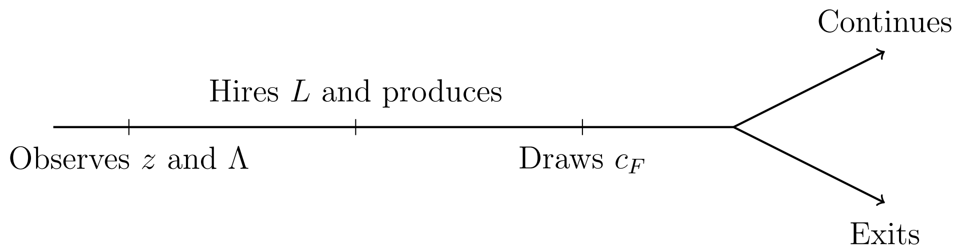
$$\Gamma(s, x) = \int_x^{\infty} t^{s-1} \epsilon^{-t} dt \tag{3.8}$$

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. Similar forces exist in models of oligopolistic competition with a finite number of firms, such as Atkeson and Burstein (2008). However, this specification accommodates a continuum of firms 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. 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 $-\epsilon/\sigma$.

Figure 5: Timing for incumbent establishments



Intermediate goods producers

At each date t, a mass N_t of intermediate goods producers use labor to produce differentiated goods. 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.

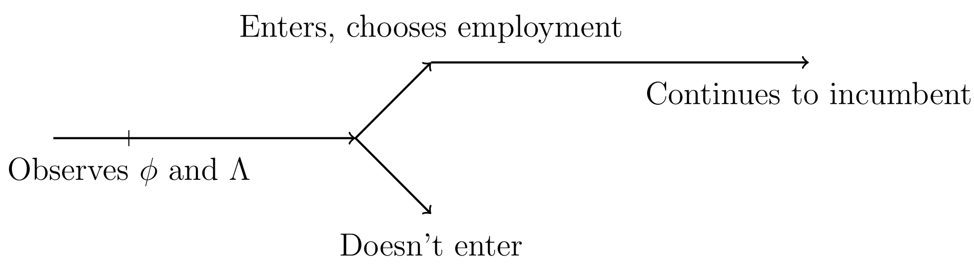
Each period, each producer observes its idiosyncratic productivity z and the state of the aggregate economy, Λ. It then hires workers, produces output, and sells its differentiated variety to the final goods producer. Producers face labor adjustment costs $\phi(L, L')$ as a function of last period's employment L and their current employment L'. After selling their output and paying adjustment cost, each producer the 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.[3]The information structure and timing are summarized in Figure 5.

Let Λ summarize aggregate states that are relevant to each producer. The recursive

[3]In the deterministic steady state, the firm 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 firms discount future streams of profits using the risk neutral discount factor β. This is equivalent to assuming either (1) this is a small open economy and the interest rate is fixed or (2) all firms are owned by a measure zero, risk-neutral mutual fund who distributes profits to the households. The reason that I choose a risk-neutral discount rate is that the preference specification I use down 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 his, the interest rate affects firm dynamics. To avoid mischaracterizing the impact of falling entry on aggregate employment, I fix the discount rate and thus the interest rate.

In Appendix F, I study the response of the economy to aggregate shocks when firms 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 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 firms more hesitant to hire.

Figure 6: Timing for potential entrants



problem of an incumbent establishment who employed L employees last period, has productivity z and has paid fixed cost ϕ_F is listed below.

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

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

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

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

Entrants

Each period, a mass M_t of potential entrants considers whether to begin producing or not. Each entrant draws an idiosyncratic signal of their future productivity $\phi \sim F$ and decides whether or not 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 $H(z|\phi)$. Figure 6 summarizes the information structure for potential entrants.

The value of a potential entrant who has drawn productivity signal ϕ is:

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

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 $\hat{\phi}$.

An alternative to the selection model of entry presented here is free entry. In that model, there is an unlimited mass of potential entrants, each of whom decides whether or not to enter without having any signal about their future productivity. In Appendix G, I discuss this model and its implications for my results.

Equilibrium

A recursive stationary equilibrium is:

1. Aggregate output Y, consumption C, labor supply L, a wage W, and a demand index D,
2. Policy functions $y(z, L)$ and $L(z, L)$,
3. Entry and production decisions,
4. Value functions V and V_E, and
5. A distribution over states $\Lambda(z, \ell)$ and a mass of entrants $M > 0$.

such that

1. The firms' policy functions satisfy their recursive definitions,
2. Policy functions are optimal given value functions and aggregate quantities,
3. The labor and goods markets clear,
4. Consumption C and labor supply L satisfy the household first order condition, and
5. The stationary distribution is consistent with the exogenous law of motion of productivity, the policy functions of incumbent firms, and the mass of new producers.

Aggregation

There are useful aggregation results for this economy.[4] 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 firm–level inverse productivities:

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

[4]Note that solving the model still requires approximating the value function of the firms. See Appendix D.1 for details.

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

A key mechanism in the model is that the elasticity of demand falls with relative size, so that firms have an incentive to increase their markups as they grow relative to the market. In this section, I provide evidence that large firms increase their markups as their market shares rise. I will use the estimates of the size of this relationship to calibrate the quantitative model.

Motivating Empirical framework

I motivate the empirical framework I use with the production function approach popularized recently by De Loecker and Warzynski (2012). Consider a firm with a production function in a variable input L and a static input K.[5] The distinction between variable and static inputs is that the firm can costlessly adjust its variable input use, while its static inputs may be subject to adjustment costs. The ability of the firm to produce might depend on conditions out of the firm's control, such as productivity, which I summarize with A. The production function can be expressed as:

$$Y = Q(A; K, L) \tag{4.1}$$

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

[5] 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.

$$WL = \alpha \frac{PY}{\mu} \tag{4.2}$$

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:

$$\log WL = \log \alpha + \log PY - \log \mu \tag{4.3}$$

Consider the following regression for firm f in year t:

$$\log WL_{f,t} = \tilde{\alpha}_{f,t} + \beta \log P_{f,t}Y_{f,t} + \epsilon_{f,t} \tag{4.4}$$

If the output elasticity α does not vary with output, then an expression for the regression coefficient β is:

$$\beta = 1 - \frac{\mathrm{Cov}(\log PY, \log \mu)}{\mathrm{Var}(\log PY)}. \tag{4.5}$$

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.

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 1985-2018, exclude financial firms and utilities, and for my baseline results classify firms using the Fama-French-49 industry definition.[6]

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% of firms in the US, but the sum of their sales is around 75% of nominal gross national income and their total employment accounts for 30% of nonfarm payroll. Table 2 shows several statistics for a few variables in the Compustat sample. The average firm has 6,800 employees, $875 Million in cost of goods sold (COGS), and $1.274 Billion in sales. The firm size distribution is heavily right skewed; for example, while the mean firm has 6800 employees, the median firm only has 700. Similarly, the median values of COGS and sales are each at least an order of magnitude smaller than their means.

[6] 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 similar results hold using SIC and NAICS definitions at various levels of granularity.

Table 2: Summary statistics of several Compustat variables

Variable	Mean	Median	25th Pct	75th Pct	Std. Dev.
Employment (1000s)	6.814	0.700	0.131	3.414	32.419
COGS ($ Millions)	874.1	48.7	9.2	271.7	5846
Sales ($ Millions)	1274	77.5	14.6	429.9	7858
Sales/COGS	2.298	1.457	1.243	1.897	23

The markup-market share relationship

To quantify how much firms increase their markups when their market shares rise, I estimate the following regression:

$$\log(WL)_{ift} = \alpha_{g(ift)} + \beta \log(PY)_{ift} + \epsilon_{ift} \tag{4.6}$$

where ift denotes the observation for firm f in industry i at date t. I estimate a variety of specifications for the variable cost WL and choices of fixed effects $g(ift)$. Table 3 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 of the number of observations of COGS and EMP in the dataset.

Table 3: 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	Log difference
Fixed Effects	Industry × Year	Firm + Industry × Year	Industry × Year

Consistent with the hypothesis that firms increase their markups as their market shares grow, the estimated regression coefficient is statistically less than one across all nine specifications. My preferred specification is (3). In column (3), I estimate the regression using one-year growth rates.[7] 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 cost of goods sold. 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.

Column (1) depicts the results of the regressions using industry-year fixed effects. If we interpret these regressions as the within-firm elasticity of variable input use to revenue, the implicit assumption in column (1) is that all firms within each industry in each year share the same output elasticity α. The numbers reported are interpretable as the difference in variable input use when comparing two firms within an industry relative to their difference in sales. The estimated coefficients in this specification are much closer to 1 than in specifications (2) and (3). This suggests that there might be permanent differences between firms that drive their differential variable input

[7] 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 4: Markups and revenue, Production Function Approach Interpretation

Variable cost measure	$\partial\mu/\partial \log PY$		
	(1)	(2)	(3)
$\log EMP$	0.1616	0.3735	0.644
	(0.0009***)	(0.0016***)	(0.0137***)
$\log XLR$	0.1017	0.3284	0.5737
	(0.003***)	(0.007***)	(0.007***)
$\log COGS$	0.0737	0.217	0.346
	(0.0007***)	(0.002***)	(0.002***)

use: firms with high relative sales may have more variable input intensive production technologies.

The fixed effects in column (1) absorb any variation in the elasticity of output parameter, α, that is common to all firms within an industry. In columns (2) and (3), I control for firm heterogeneity, allowing production functions to vary at a finer level. In column (2), production functions are allowed to have a fixed firm component α_f plus a time–varying industry component $\alpha_{i,t}$. In column (3), which uses log-differences, I assume that the output elasticity must change at the same rate for every firm within an industry from year-to-year.

Production Function Approach Interpretation

In the static framework I discussed at the beginning of this section, a coefficient less than 1 is consistent with markups that rise with relative sales. We can quantify the relationship between log markups μ and revenue by the complement to the regression coefficient estimated above.

Table 4 summarizes this structural interpretation. The most conservative estimate relies on specification (1) and uses cost of goods sold as the measure of variable input cost. It implies that in the average industry, a firm with 1 percent higher sales has markups that are 7 basis points higher. Specifications (2) and (3) allow for heterogeneity in production functions within industry and imply that markups increase by more. Specification (3) using COGS, for example, states that when a firms' sales grow at a rate 1 percent above the industry average, it increases its markup by 35 basis points. The difference in these regression coefficients shows that it is important to control for firm heterogeneity when estimating the relationship between markups and size.

Markups vs. Production Function

In interpreting these regression coefficients, I allow for a variety of specifications to account for production function heterogeneity. However, across all specifications, I assume that the output elasticity does not vary with revenue PY. This holds clearly in the case of Cobb-Douglas, but is not generally true. If, for example, $\log \alpha$ decreases with output, then the deviation of $\hat{\beta}$ from 1 could be attributed to production function variation rather than to markup variation.

To investigate this hypothesis, I ask whether capital/labor ratios vary with firm size among Compustat firms. I use $PPEGT$ and $PPENT$ as measures of the capital stock. I estimate

$$\frac{K_{ift}}{L_{ift}} = \alpha_{it} + \beta P_{ift} Y_{ift} + \epsilon_{ift} \tag{4.7}$$

Across both specifications for the capital stock, the estimated β coefficient is not statistically different from 0. While there may be shortcomings in the measurement of capital in Compustat, a regression of the capital stock directly on revenue reveals regression coefficients of nearly 1. If labor intensity fell with firm size, we would expect capital-labor ratios to rise with firm size. So, the lack of variation in capital-labor ratios with revenue suggests that it is not production function variation that pushes $\beta < 1$.

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 (i.e., it was not truly variable), then the static first order condition in the production function approach would not hold. In that case, the quantity μ represents any wedge distorting the firms' production choices away from their static optima.

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, it 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 jointly estimate 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.

Relationship to De Loecker and Eeckhout (2017)

De Loecker and Eeckhout (2017) also use the production function approach to study markups. The key difference between my approach and theirs is that my focus is on how markups vary within firms over time, while theirs is on estimating the average level of markups. Because I am interested in how markups vary within firms rather than in their average level, I do not estimate α directly. Instead, I allow fixed effects to absorb any variation in α across industries or over time.

This approach avoids two issues with the standard approach. First, not estimating α avoids the issue of how to compute quantity in Compustat. In De Loecker and Eeckhout (2017), estimating α requires a measure of real output for each firm. To obtain this measure, they deflate each firms' sales by an industry deflator to compute quantity. However, if firms within an industry set different prices, as is true in the model I use later, this is a problematic assumption. Bond et al. (2020) formally discuss the shortcomings of this approach.

Second, not estimating the output elasticity directly allows for more heterogeneity across firms. De Loecker and Eeckhout (2017) assume that the elasticity of output α is common to all firms within a given industry in a given year. This is a necessary assumption to be able to precisely estimate this parameter. However, in my specification, because $\log \alpha$ is additive in the estimation equation, it is swept out by fixed effects. So, I show regressions in which firms share production functions within an industry, but I also discuss specifications in which α varies across firms within an industry–year. The latter estimates imply that markups vary more strongly with market share than the estimates from De Loecker and Eeckhout (2017) would imply.

The rise in the markup-revenue relationship

I have shown evidence that markups covary positively with market share in a panel of large firms. As I show in this section, this relationship has grown stronger over the past 30 years.

Figure 7 summarizes the results of estimating each of the 9 regression specifications of variable input use on relative sales as before, using centered rolling 5-year windows. For both employment and cost of goods sold, the coefficients decline by significant amounts from 1985 to 2015. The plots using XLR exhibit noisier estimates but still

Table 5: 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

generally decline after 2000.[8] Table 5 summarizes the endpoint estimates for each of the specifications. Across all specifications, the elasticity of variable input costs to revenue declined over the sample.

The most conservative estimate, using cost of goods sold and only within-industry between-firm variation suggests that markups used to increase by only 3 basis points for every 1 percent increase in sales and now increase by 10 basis points for the same increase in sales. Controlling for heterogeneity across firms increases both the initial level and the slope of its secular trend. Using the cost of goods sold and specification (3) implies markup elasticities to relative sales of 20% in 1990 and 55% in 2015. Using employment or the wage bill as the measure increases the end–of–sample estimate to 75%. All of these estimates imply that large firms increase their markups more strongly as their market shares grow today relative to 1985.

[8]The larger standard errors and wider fluctuations are not surprising given the sparsity of data available for that measure.

Figure 7: Variable Input–Revenue Relationship, Rolling Windows

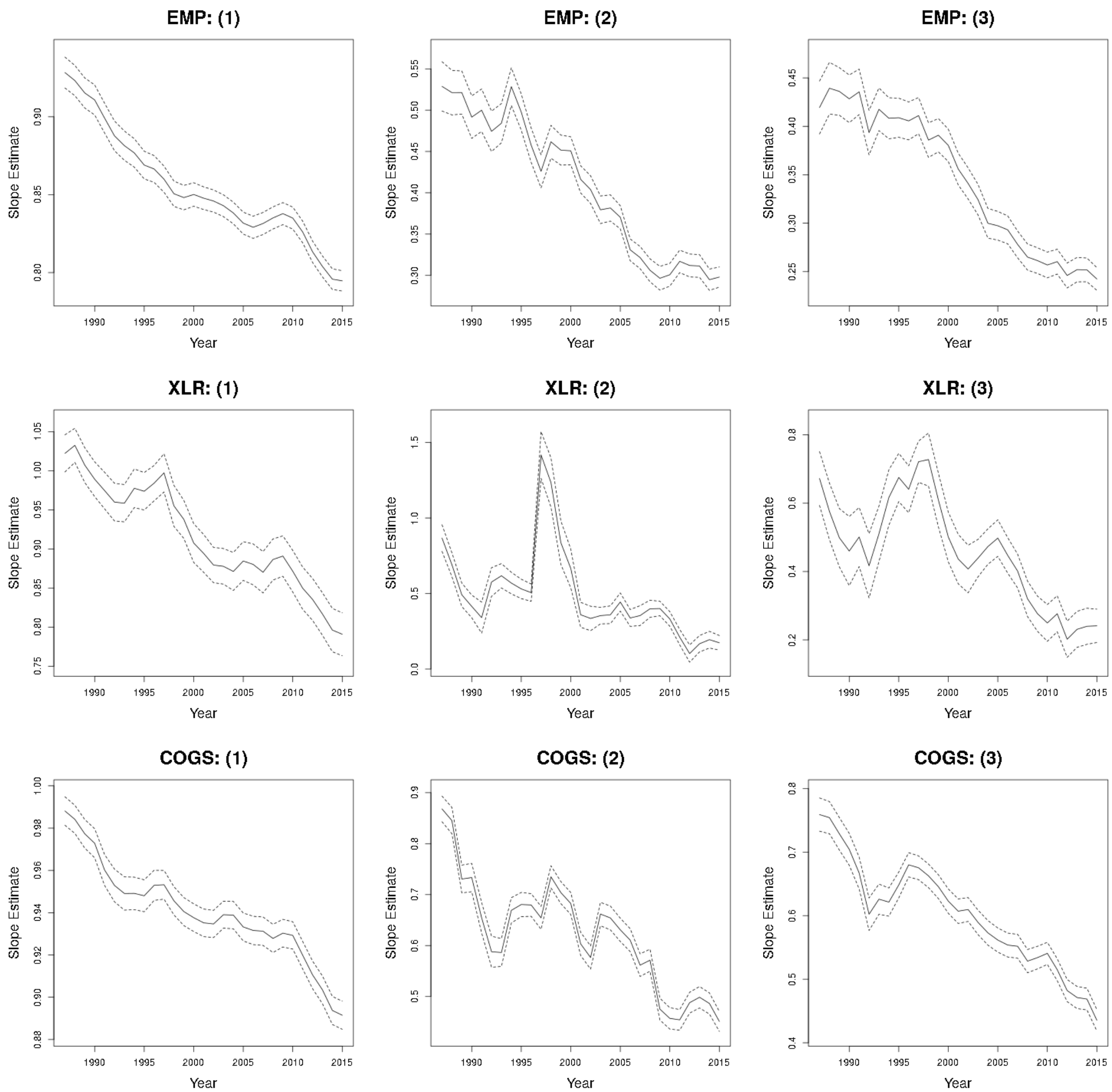


Table 6: Reallocation in Compustat, 2010

Measure	Reallocation
EMP	6.17 %
XLR	7.24 %
$SALE$	14.15 %

Markups and labor reallocation

In this section, I show that a stronger relationship between markups and market share is consistent with the fall in labor reallocation documented by Decker et al. (2018). Taking first difference of the the first order condition of the firm discussed earlier gives a decomposition of the cross–sectional variance of sales growth ("sales reallocation") into the variance of employment growth ("employment reallocation") and two terms about markup variation:

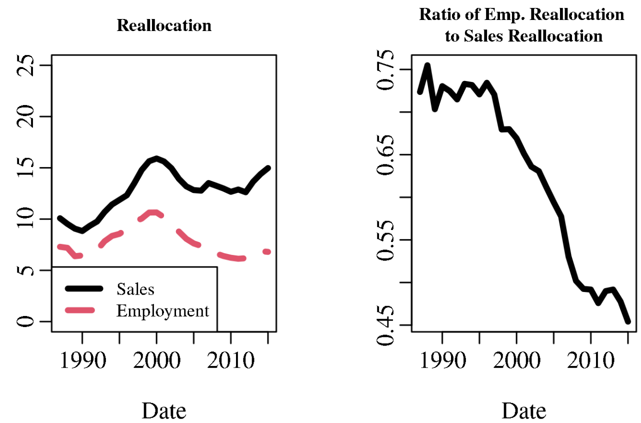
$$\underbrace{\text{Var}(\Delta \log PY)}_{\text{Sales reallocation}} = \underbrace{\text{Var}(\Delta \log WL)}_{\text{Employment reallocation}} + \underbrace{\text{Var}(\Delta \log \mu) + 2\text{Cov}(\Delta \log \mu, \Delta \log L)}_{\text{Markup variation}} \quad (4.8)$$

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 and more variation in markup growth implies a wedge between sales and employment reallocation, and so the higher correlation between markups and firm size that I document could drive a decline in employment reallocation relative to sales reallocation.

Table 6 summarizes these measures in 2010. As it shows, employment and wage bill reallocation are roughly half the size of COGS and revenue dynamism. The difference implies that about half of sales reallocation is due to the dispersion in markup growth and its covariance with employment growth.

These measures have not been stable over time. As emphasized in Decker et al. (2018), employment reallocation rose during the 1990s and then fell again. The red line in Figure 8 confirms that these patterns hold in Compustat. A less–studied fact is that sales reallocation rose during the 1990s but has remained stable since then, implying that the wedge between the two measures has widened since 1995. The right panel shows the ratio of labor reallocation to sales reallocation over the same period. While employment reallocation used to be around 75% of sales reallocation, it has fallen to 45%.

Figure 8: Employment and Sales Dynamism



The fall in input reallocation relative to sales reallocation implies that the "markup variation" term has risen. Fact 2 suggests that part of this increase is due to a rise in the covariance between markups and employment. Later in the paper, I use the structural model to show that an increase in the superelasticity of demand can quantitatively account for this rising wedge.

Summary

In this section, I show three facts in a panel of firms from 1985 to the present. First, I show that variable input use varies less than one–for–one at the firm level. This holds across a variety of measures of variable input use. Second, input use elasticity with respect to revenue has declined consistently and dramatically since 1985. Third, I show that the cross–sectional variance of within–firm employment growth (employment reallocation) has fallen relative to sales dynamism.

Under the assumption of no adjustment costs on variable inputs, the first fact implies that markups rise with firm relative size. I later allow for adjustment costs, estimating a structural model featuring both adjustment costs and markups that systematically vary with market share. I use external data on the size of adjustment costs to discipline the adjustment cost channel, finding that the market power story is quite strong.

At the end of the paper, I revisit the secular trends in the markup–size relationship and the wedge between labor and sales reallocation. I show that one structural change can account for both of these trends, and I then show that this structural change implies that cyclical variation in entry matters more for aggregate employment today than it did in 1985.

5 Steady state

In the steady state of the model, firms are heterogeneous along a number of dimensions. Each firm's idiosyncratic state variables are its productivity and employment. Firms have a lifecycle, beginning small and slowly hiring workers and becoming more productive. Moreover, firms face labor adjustment costs, and so firms' output and pricing decisions are history dependent. And, firms differ in the elasticity of demand they face and thus in the markups they set.

The employment-sales regression

As I showed in Section 4, large firms increase their variable input use less than one-for-one with revenue, which suggests that their markups increase with their market share. The model can reproduce this pattern through two mechanisms: (1) the elasticity of demand falls with firm size, leading firms to increase their markups as they grow, and (2) adjustment costs prevent firms from adjusting their variable input use in response to productivity shock.

To understand the role of the superelasticity, consider the model without adjustment costs. In that case, $\phi_L = 0$, and the establishment's only idiosyncratic state variable is its productivity. As a firm's productivity rises, it produces more and its elasticity of demand falls. In response, it increases its markups. The increase in markups means that the firm increases its employment less than one-for-one with its sales. Figure 9 depicts the relationship between sales and employment in this model in blue, and the same relationship in a model with constant markups in the black dashed line. Establishments 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 firms increase their markups more with sales than small firms 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 presents a challenge in calibrating the model, since the average Compustat firm is larger than the average firm in the economy, which might lead me to overstate the extent to which markups rise with market share for the average firm. To calibrate the model, I estimate Equation (4.4) on a sample of the largest firms in the model.

The sample I use in Compustat covers about 1% of firms and 30% of U.S. non–farm payroll. In my simulated method of moments estimation procedure, I simulate a sample of firms in the model and then estimate the regression on a subsample of the top 1% of firms by sales in the model economy. This procedure generates a comparable subsample

Figure 9: Employment and sales in the frictionless model

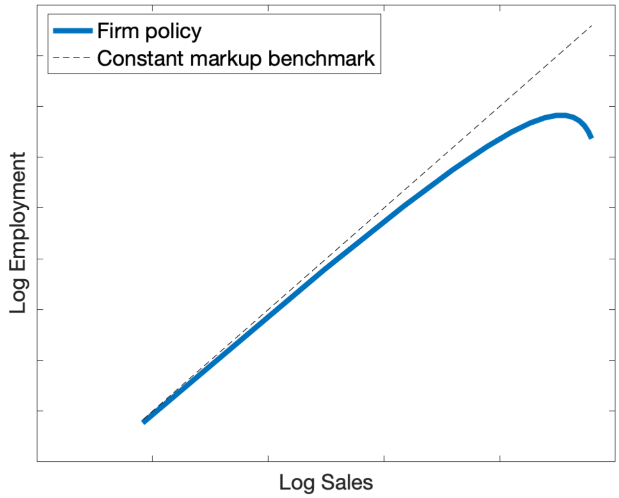


Figure 10: Identification of the superelasticity

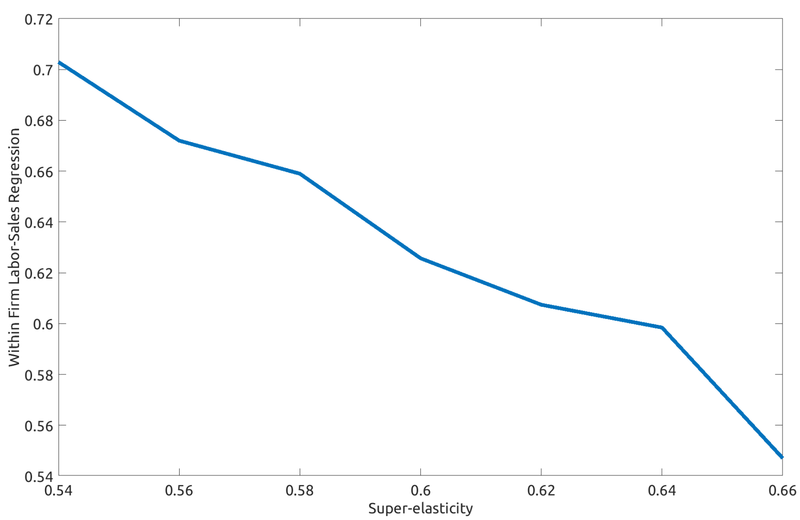
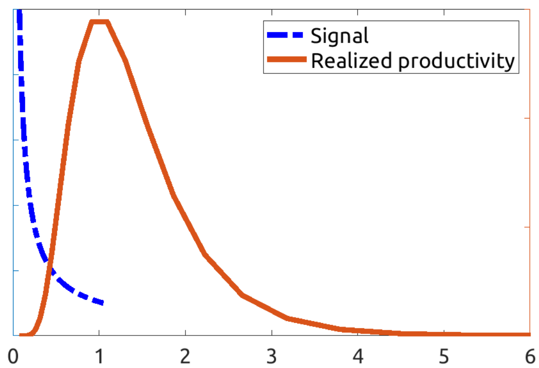


Figure 11: The distribution of the signal and productivity



to estimate the super-elasticity. In Figure 10, I plot the regression coefficient in the model at different values for the super-elasticity. As it shows, a higher super-elasticity means that large firms increase their markups more with their market shares and so they hire fewer workers in response to increases in productivity.

Calibration

Functional forms

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

$$u(C_t, L_t) = \frac{1}{1-\gamma}\left(C_t - \psi \frac{L_t^{1+\nu}}{1+\nu}\right)^{1-\gamma} \tag{5.1}$$

These preferences imply a labor supply curve:

$$\psi L_t^{\nu} = W_t \tag{5.2}$$

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 productivity distribution for entrants is equal to a shifted version of the log normal stationary distribution of productivity implied by the AR(1) process for productivity. Figure 11 depicts the distribution of new entrants and this stationary distribution.

Table 7: 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
γ	Exogenous exit rate	1.5%	
M	Mass of entrants	1	Normalization
ν	Inverse Frisch Elasticity	0.5	
δ	Job separation rate	0.19	

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 are summarized in Table 7. 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 correspond to one or two moments. Below, I provide intuition for the calibration strategy. The persistence of productivity and dispersion in its innovations affect the cross–sectional variance of firm–level log sales growth and the share of sales among the 10% largest firms. The size of the shift in log productivity affects the relative size of entering firms. I identify the degree of adjustment costs with the auto-correlation of firm-level log employment growth, which I estimate to be 12.81% 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 super–elasticity, on the other hand, affects the relationship between firm size and the markup and so affects the within–firm regression coefficient of employment on sales. For the baseline calibration, I use an estimate of 0.57, which matches the coefficient using specification (3) and COGS as the measure of variable cost. Table 8 summarizes the parameter choices as well as their identifying moments in the model and in the data.

The model performs well along a number of targeted and untargeted moments. Figure 9 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

Table 8: Calibrated parameters

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

Table 9: Calibration Targets & Model Fit

Moment	Target	Source	Model moment
$\text{Var}(\Delta \log L))$	6.17%	Compustat	5.8%
$\text{Var}(\Delta \log PY))$	14.15%	Compustat	13.4%
$\rho(\Delta \log L_t, \Delta \log L_{t-1})$	0.13	Compustat	0.1281
$\rho(\Delta \log P_t Y_t, \Delta \log P_{t-1} Y_{t-1})$	0.12	Compustat	0.122
Labor–sales regression	0.654	Compustat	0.628
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	1.264
Top 10% share of sales	75%	Compustat, industry average	69%
Share of employment at young firms	30%	BDS	32.97%

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

young establishments that I estimate in the BDS. Fitting these are 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 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 is likely due to the long right tail of sales in the data that is not present in a model with log-normal productivity.

Superelasticity estimate

My estimate of the superelasticity is consistent with estimates from a broad literature that uses firm–level data. As summarized in Table 10, estimates of the superelasticity using microdata tend to be below 1. My estimates are closest to Amiti, Itskhoki and Konings (2019), Berger and Vavra (2019), and Gopinath, Itskhoki and Rigobon

Table 10: Selected parameterizations of the superelasticity

Paper	ϵ/σ	Note
This paper	0.60	
Edmond, Midrigan and Xu (2018)	0.14	
Amiti, Itskhoki and Konings (2019)	0.26	
Berger and Vavra (2019)	0.47	
Gopinath, Itskhoki and Rigobon (2010)	0.6	
Goldberg and Hellerstein (2013)	0.8	Estimate for beer
Nakamura and Zerom (2010)	4.6	Estimate for coffee
Lindé and Trabandt (2019)	10	
Smets and Wouters (2007)	12.55	

Estimates below horizontal line are based on macro data, above line are based on micro-data

(2010), who estimate the superelasticity using within-firm price responses to marginal cost shocks.

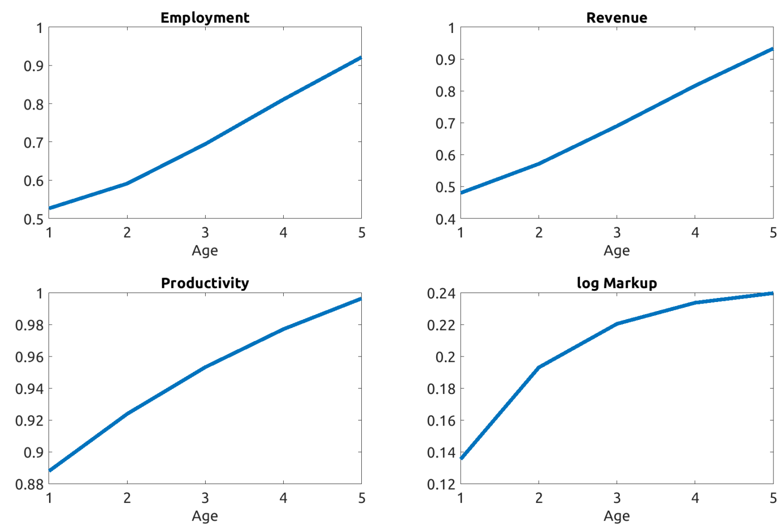
Edmond, Midrigan and Xu (2018) estimate the superelasticity using a cross-sectional regression of a transformation of the markup, estimated following De Loecker and Eeckhout (2017), on relative sales. I find a somewhat larger estimate of the super-elasticity than they do. As I discussed before, following De Loecker and Eeckhout (2017) requires assuming that firms within an industry all share the same production function. I find that regressions that relax this assumption imply that markups covary much more with market share.

Consistent with other studies that use micro-data 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 vs. labor adjustment

As discussed before, the within-firm regression coefficient of employment growth on sales growth could be less than one for several reasons. In the model, the two forces that generate the less-than-one-for-one regression coefficient are the positive super-elasticity of demand and labor adjustment costs. The model allows me to decompose the reduced-form regression coefficient into each component. Recall that the regression coefficient in the model is 0.628. When I set $\phi_L = 0$, re-solve the model, simulate a panel of firms

Figure 12: The lifecycle of the firm in the quantitative model



in the new model, and estimate the regression coefficient, I find $\hat{\beta}_L = 0.65$. When I set the super-elasticity 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% and 20% of the deviation of the regression coefficient from 1.

Aggregate parameters

There are some parameters whose values do not affect the steady state of the economy, only its response to aggregate shocks. These are the inverse Frisch elasticity, which I set to be $\nu = 1/2$, following Clementi and Palazzo (2016), and the disutility of labor parameter, ψ, which I set so that the steady state wage is 1.

The lifecycle of the firm

Firms in the model, as in the data, begin small and grow slowly. Figure 12 shows that the average entering producer employs around 50% of the labor force of the average incumbent firm. They reach the size of the average firm by around age 5. The model achieves this in two ways: (1) the average productivity of entering firms is lower than that of incumbents and slowly reverts to the mean and (2) labor adjustment costs further slow the growth of new firms.

Firms' markups in the model also follow a lifecycle pattern, beginning low and slowly increasing. The desire to set high markups derives from a demand elasticity that decreases with relative size. Since young firms' market shares slowly grow, their

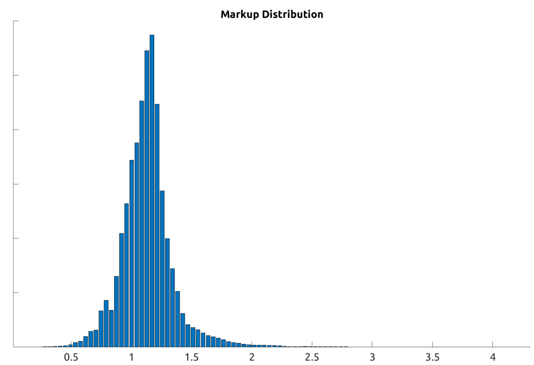


Figure 13: The cost-weighted distribution of markups

markups also slowly increase with age. The cost–weighted average markup increases by around 10 percentage points over the first 5 years of a firms' life in the model.

The distribution of markups

Firms in the steady state of the model set heterogeneous markups. Consistent with recent evidence on markups (see Edmond, Midrigan and Xu (2018) and De Loecker and Eeckhout (2017)), the cost-weighted average markup in the model is around 1.25. The sales-weighted average in the model is 1.285, which is far below the value that De Loecker and Eeckhout (2017) estimate. The cost-weighted average markup is the inverse of the labor share, and so is the relevant measure in this model.

Figure 13 depicts the employment–weighted distribution of markups in the model. Most firms set markups between 1 and 2. Some set markups below 1, reflecting labor adjustment costs. There are a few large firms who set markups above 2, and those firms tend to be large, both in terms of sales and employment.

The non–degenerate distribution of markups is novel relative to the literature on entry over the business cycle. While Jaimovich and Floetotto (2008) and Bilbiie, Ghironi and Melitz (2012) study variation in markups in response to entry, they solve for symmetric equilibria in which all firms set the same markup and entering firms are the same size as incumbents. The distribution of markups is also an innovation realative to Siemer (2014), Moreira (2017), and Clementi and Palazzo (2016), who all study models in which entrants are smaller than incumbents and firms face heterogeneous productivities. However, their models do not imply markups that systematically vary with market share. As I show later, these models understate the effects of entry on aggregate employment.

6 Shocks to entry over the business cycle

To study business cycle fluctuations in entry, I solve for the response of the model economy to a one-time unexpected shock to the mass of potential entrants. 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 D.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 (2014) or a fall in demand, as in Moreira (2016).

An entry shock

Figure 14 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. Since the labor share is the inverse of the average markup, it falls by 80 basis points. Effective TFP, equal to the ratio of output to aggregate employment, falls gradually by nearly one percent. Employment falls by 2 percent on impact, and output falls by a bit over 2 percent. The wage satisfies the household labor supply equation and falls by around 1 percent.

In response to the shock, the entry rate and share of employment among entrants and young firms fall. Figure 15 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%.

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 aggregate model is summarized by an aggregate production function (Equation 6.1), the definition of the markup as the inverse labor share (Equation 6.2), and the labor supply equation (Equation 6.3).

$$Y_t = Z_t L_t \tag{6.1}$$

Figure 14: The response of the baseline quantitative model to an MIT shock

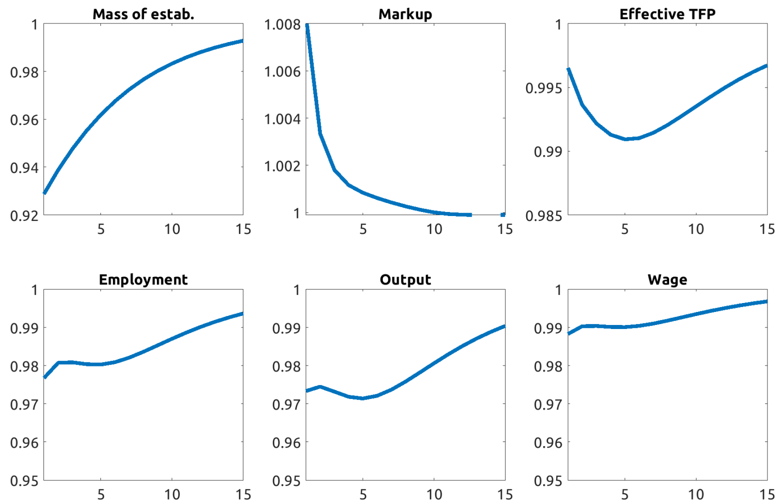


Figure 15: Entrants following the shock

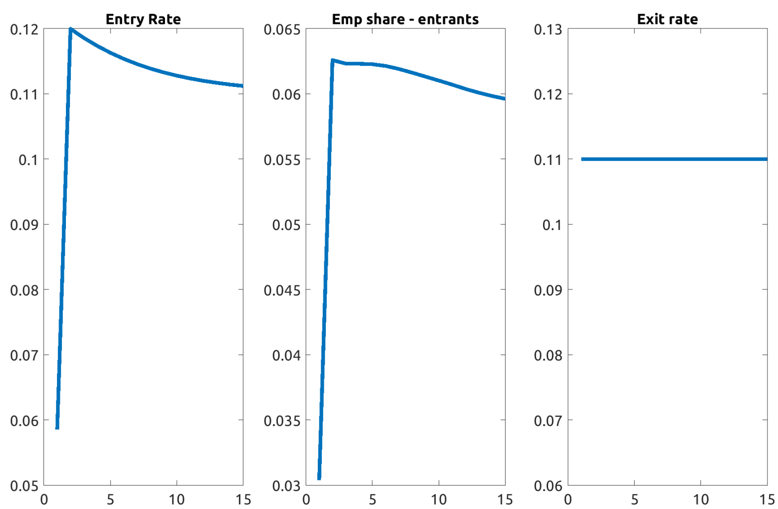
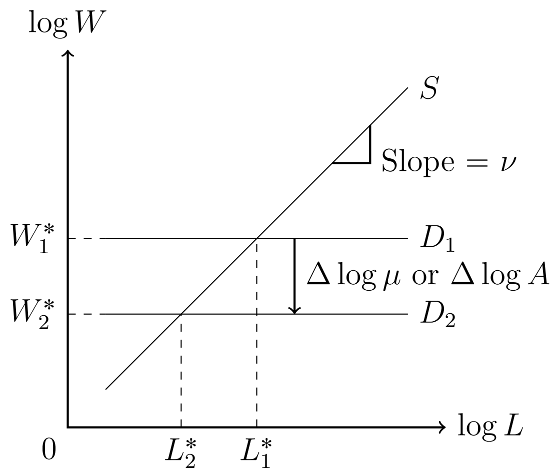


Figure 16: A rise in the markup or a fall in effective TFP



$$\mu_t = \frac{Y_t}{W_t L_t} \tag{6.2}$$

$$W_t = \psi L_t^{\nu} \tag{6.3}$$

Given paths for the cost–weighted markup μ_t and aggregate effective productivity A_t, equations (6.1 - 6.3) 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.

How much of a fall in employment the rise in the markup and the fall in TFP each causes is easy to read off of a simple supply–demand diagram. Some algebra shows that Equations (6.1 - 6.3) can be expressed as labor supply and labor demand equations:

$$\log W_t = \log \psi + \nu \log L_t \tag{6.4}$$

$$\log W_t = \log A_t - \log \mu_t \tag{6.5}$$

A rise in the markup (or a fall in TFP) shifts labor demand down and causes the wage to fall by $\Delta \log \mu$ (or fall, in the case of TFP, by $\Delta \log TFP$) and employment to fall by $(1/\nu) \times \Delta \log \mu$. Since $\nu = 0.5$, the decline in employment is double the rise in the markup (or the fall in TFP). Figure 16 depicts this graphically. A rise in the markup or a fall in effective TFP leads the demand curve to shift down. The slope of the labor supply curve (ν) determines how much this shift in demand leads to a fall in employment and the wage.

Figure 17: Decomposition of entry shock

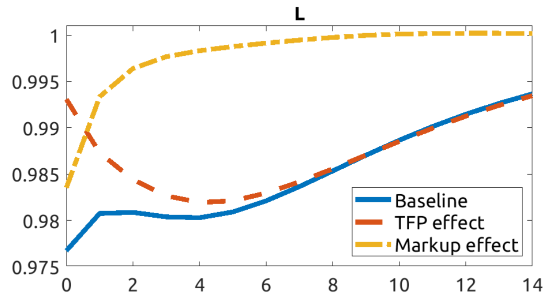


Figure 17 depicts the paths of output, employment, and the wage under different paths for the markup and productivity. In blue, I allow both to follow their equilibrium paths. In red, I hold the markup fixed, and in yellow, I hold TFP fixed. As they show, the rising markup generates a fall of 1.5% in 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.

Aggregate TFP

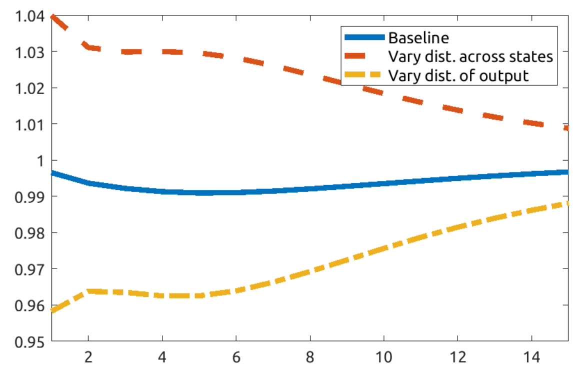
The decline in aggregate TFP accounts for about a quarter of the contraction in employment on impact. To understand why aggregate TFP falls, I decompose its fluctuations into movements due to the change in the distribution of firms and those due to the change in the allocation of output across firms. Recall the definition of Z_t:

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

As this definition shows, Z_t might fluctuate because of changes in $q_t(z, L)$ or changes in $d\Lambda_t(z, L)$. Figure 18 decomposes the path of TFP into each of these two changes. The red dashed line holds fixed the function $q_t(z, L) = q_{SS}(z, L)$ but allows the distribution $d\Lambda_t(z, L)$ to vary. TFP in this exercise rises because entrants are less productive than incumbents, and so a fall in entry leaves the economy with fewer unproductive producers.

The yellow dot-dashed line shows the path of TFP, holding fixed $\Lambda_t(z, L) = \Lambda_{SS}(z, L)$. In this exercise, TFP falls by more than in the actual equilibrium response. This is due to the reduction in relative sales among productive establishments. So, economy-wide productivity falls because large, productive establishments raise their markups and produce less, relative to small firms, in response to the fall in entry.

Figure 18: A TFP decomposition



The cost weighted markup

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

$$\mathcal{M}_t = \int\int \mu_t(z,L)\frac{\ell_t(z,L)}{L_t}d\Lambda_t(z,L) \tag{6.6}$$

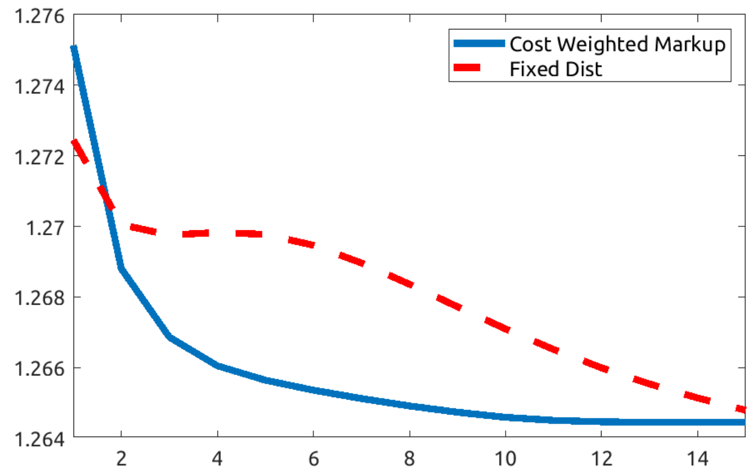
The shock to entry affects the markups of individual firms $\mu_t(z)$ and the distribution of employment across firms. Two opposing forces affect the cost–weighted markup: (1) large firms raise their markups in response to the fall in entry and (2) there is a reallocation of output away from high markup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

Adjustment costs slow the reallocation to low-markup firms. One way to see this is to compare the path of the markup holding $\ell_t(z,L)/L \times d\Lambda_t(z,L)$ fixed. Figure 19 depicts this comparison. In red, I allow markups to vary but hold the distribution of employment fixed. This plot shows that the average firm 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. Evidently, following the shock, there is reallocation of employment to small, low-markup, firms following the shock.

The role of adjustment costs

Adjustment costs slow the reallocation of output to low-markup firms. To quantify this mechanism, I compare the baseline response to the impulse response to the same shock in an economy without adjustment costs. In Figure 20, I plot the path of the cost-weighted average markup in each economy. As it shows, without adjustment costs,

Figure 19: Decomposition of the response of markups



the markup rises by 75% less than with adjustment costs.

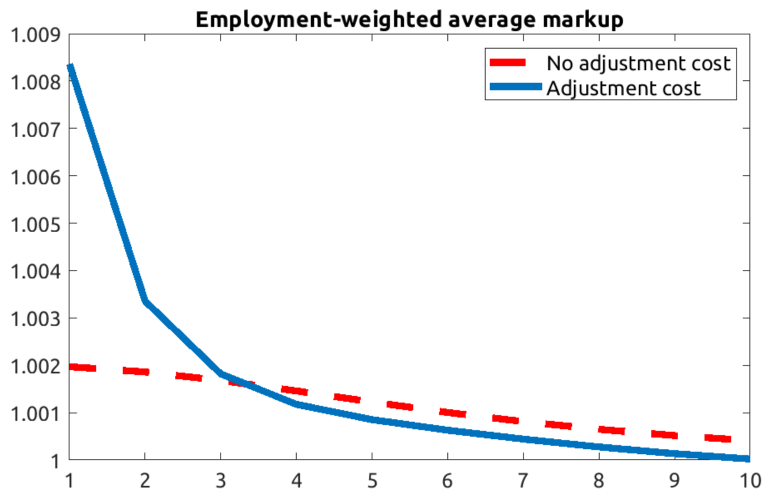
Why does employment reallocate towards low-markup firms in response to the shock? The cause of variable markups in the model is a variable elasticity of demand; small firms set lower markups because they face a higher elasticity of demand than large firms. This feature also means that small firms are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

Relationship to Arkolakis et al. (2019) and Edmond, Midrigan and Xu (2018)

Arkolakis et al. (2019) show that in a class of trade models with Pareto-distributed productivity, variable markups, no adjustment costs on variable inputs, and a choke price, there are no effects of entry on the aggregate markup. In fact, they show that there is no effect at all on the distribution of markups. My model does not satisfy the assumptions underlying that result: productivity is not Pareto distributed, there are adjustment costs, and there is no choke price. Adjustment costs, as I discussed, inhibit most of the reallocation effect. The distributional assumption turns out to take care of the rest.

Entry has almost no effect on the cost–weighted markup in Edmond, Midrigan and

Figure 20: The role of the adjustment cost in reallocation



Xu (2018) either. As they discuss, this neutrality result arises because, in response to a fall in entry, small firms grow relative to large firms, a force that leads the employment-weighted average to fall.

While adjustment frictions explain much of the difference between the results of this paper and the neutrality results of Arkolakis et al. (2019) and Edmond, Midrigan and Xu (2018), the Pareto distribution also plays a role. With Pareto productivity and no adjustment costs, a fall in entry effectively scales the productivity distribution. Because of the properties of the Pareto distribution, the scaled distribution is the same Pareto distribution with a higher lower bound. Because the smallest firms do not produce much, shifting the lower bound of the productivity distribution does not change the aggregate markup very much.

This logic does not carry through with adjustment costs and log–normal productivity. Adjustment costs, as discussed above, prevent the reallocation of output to low-productivity firms. Moreover, under the log-normal assumption, a change in entry affects both the mean and variance of the distribution of markups. A fall in entry increases concentration and thus the cost–weighted markup. I explore this argument more formally in Appendix E.

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.

The general Kimball form of the final goods production function nests CES demand. In the CES case, the aggregator is

$$\Upsilon(q) = q^{\frac{\sigma-1}{\sigma}} \tag{6.7}$$

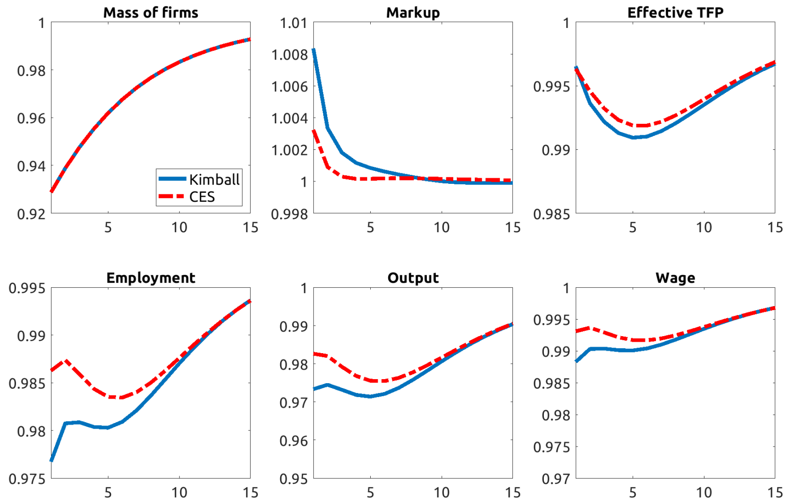
I subject each economy to the same entry shock as before. Figure 21 depicts the results of this experiment. These impulse response functions show that variable elasticity of demand generates a significant fall in employment and amplifies the effects of an entry shock. The markup rises somewhat (by 37 basis points) in the CES model because adjustment costs push firms away from their frictionless optimal solution. In response to the aggregate shock, firms have to pay an extra adjustment cost. This increases their marginal costs and in response they increase their markup. This is not measured in the model, however, and so the increased markup in the CES case reflects mis-measurement of the true marginal cost. 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 model. Because of the love-of-variety effects present in both models, effective TFP falls by a similar amount in each. However, in the Kimball model, large incumbents raise their markups in response to the increase in their market shares. This leads them to reduce their employment, causing aggregate employment to fall. The additional rise in the markup in the Kimball economy generates a nearly 75% extra fall in employment on impact in the model with variable markups. This difference disappears after around 5 years.

Shocks to the cost of entry

So far, I have shown 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 section, 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 and the markup.

Figure 22 shows the response of entry to this shock. While the entry rate falls by the same amount as in response to the shock to the mass of potential entrants, the employment share among entering firms does not move by much. The reason for this difference is that, in this model, the marginal firm considering whether or not to enter relatively unproductive and so employs few workers. A surprise increase in the cost of entry leads only these marginal firms to decide to not enter. Note that the path of the share of employment at new entrants is inconsistent with the path during the Great

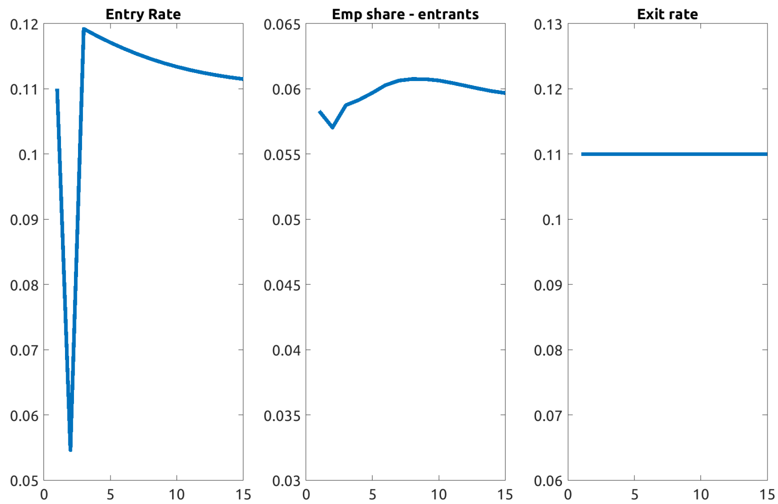
Figure 21: Entry shock in the Kimball and CES models



Recession, when the share of employment at new entrants fell significantly.

Because the entry cost shock only affects the smallest entrants in the economy, the shock has very little effect on the aggregate markup. Figure 23 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.

Figure 22: Behavior of entrants following a shock to the cost of entry



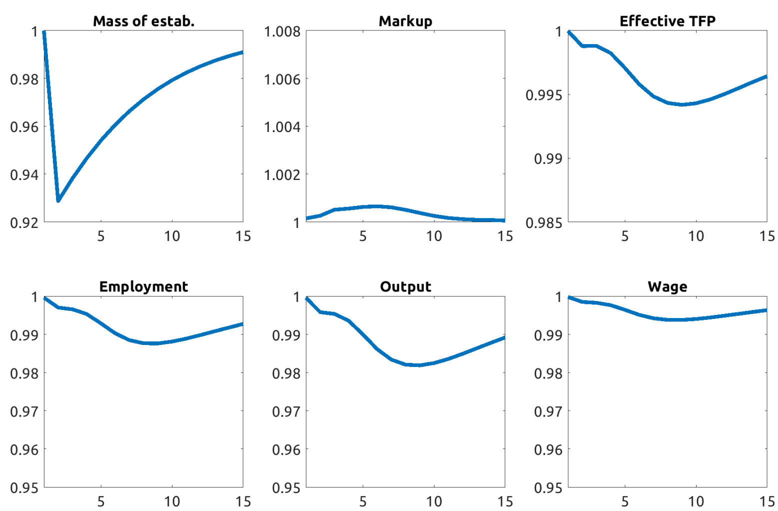


Figure 23: Entry cost shock in the Kimball model

7 Quantitative applications

In this section, I study two applications of the model. In the first, I study the role of entry in the sharp contraction in employment during the Great Recession and its subsequent slow recovery. I show that an entry shock that reproduces the path of the mass of firms during the Great Recession leads employment to fall persistently by 5 percent, returning to trend only by 2020. In the second, I return to the secular trend in the input-revenue regression coefficient that I documented in Section 4. I show that an increase in the superelasticity of demand that accounts for this trend also accounts for the fall in labor reallocation relative to sales reallocation. I then compare entry shocks in a model calibrated to 1985 data to a model calibrated to 2015 data and find that entry's effects on aggregate employment are significantly larger in the 2015 calibration.

The Great Recession

Entry during the Great Recession fell sharply and remained depressed for many years after the end of the recession. In this section, I quantify the effect of markups on employment during that episode.

Figure 24 shows the path of the number of establishments since 1977 relative to 2007. As it shows, during the Great Recession, the number of establishments fell by around 7 percent. While the number of firms typically falls in a recession, this decline was unprecedented in both size and duration.

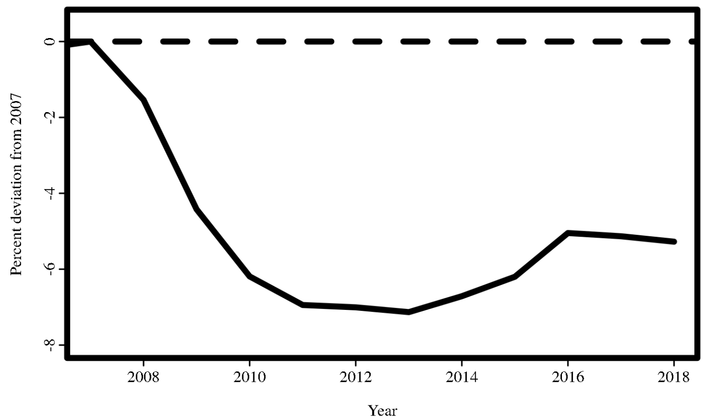
Employment among all firms fell sharply and recovered slowly during the Great Recession, but it fell especially persistently among young firms. Aggregate employment fell by 6 percent over 3 years. This headline number masks considerable heterogeneity across firms. Employment at entrants and at firms below the age of 5 fell by 30 percent and remained depressed through 2014, by which point aggregate employment had returned to its original level. The entrant establishment share of employment was around 5.5% going into the recession, and it fell to about 4% by 2012. The young firm share of employment follows a similar trajectory from slightly below 30% to nearly 20% over the same period.

The Great Recession in the model

To understand the effects of the fall in entry during the Great Recession on markups, I feed in a sequence of shocks to the mass of potential entrants so that the path of the number of establishments in the model follows its path in data from 2007 to 2014. As before, I perform this experiment in both the constant elasticity and Kimball models.

Figure 26 depicts the results of this experiment. The fall in entry leads the mass of firms to gradually fall by 6 percent. The labor share falls by nearly 80 basis points

Figure 24: Number of establishments per capita, relative to 2007



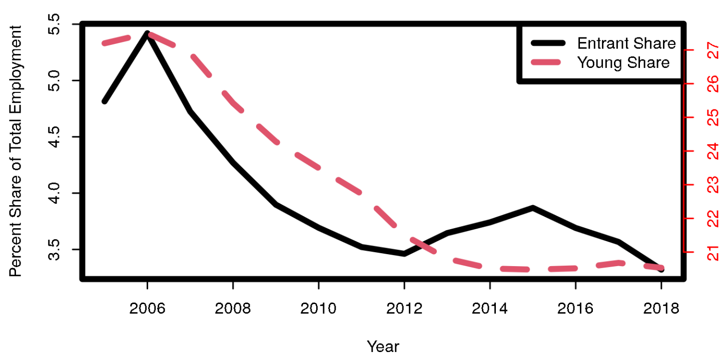
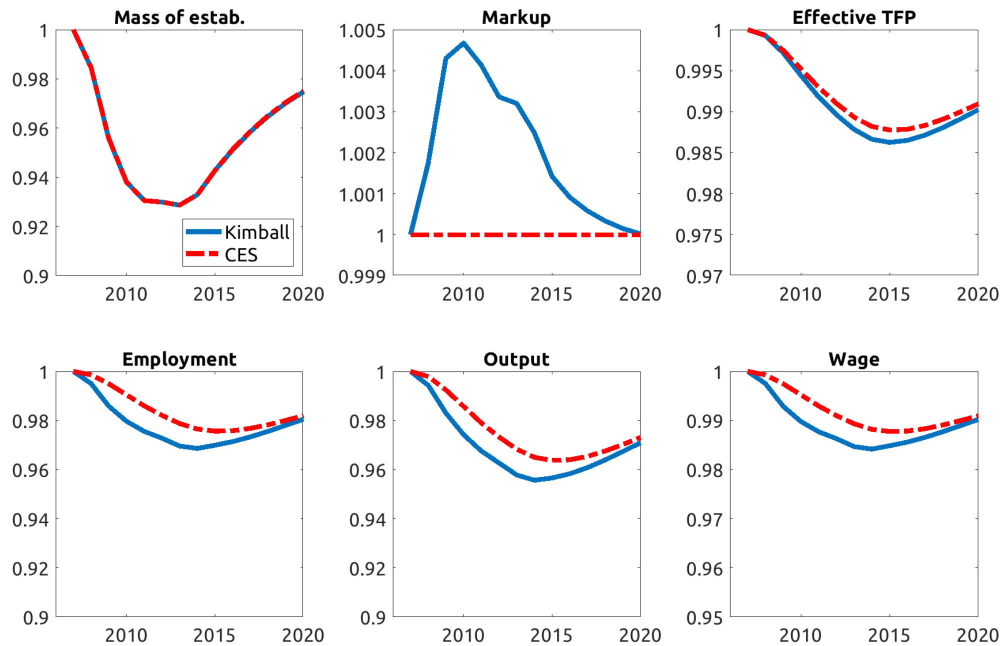


Figure 25: Age and employment

Figure 26: The Great Recession



in the Kimball model, and effective TFP falls by 2 percent. Employment falls by 5 percent and only gradually returns to its pre-recession trend in 2020. Comparing the CES impulse response functions to the Kimball ones, the variable markups channel accounts for nearly half of the fall in employment coming from the fall in entry.

The rising importance of markups for business cycles

As I showed in Section 4, the relationship between firm size and variable input use has changed dramatically over the past 30 years. In this section, I study the response of the model economy 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.

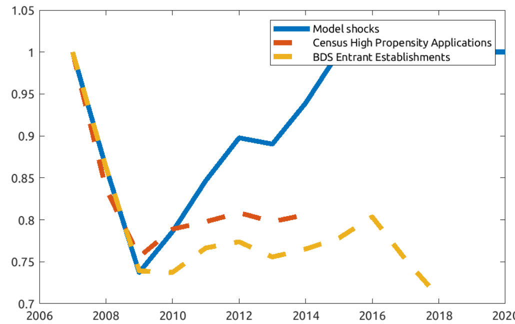


Figure 27: The sequence of shocks in the Great Recession simulation

Table 11: Selected moments, 1985 vs 2015 calibration

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

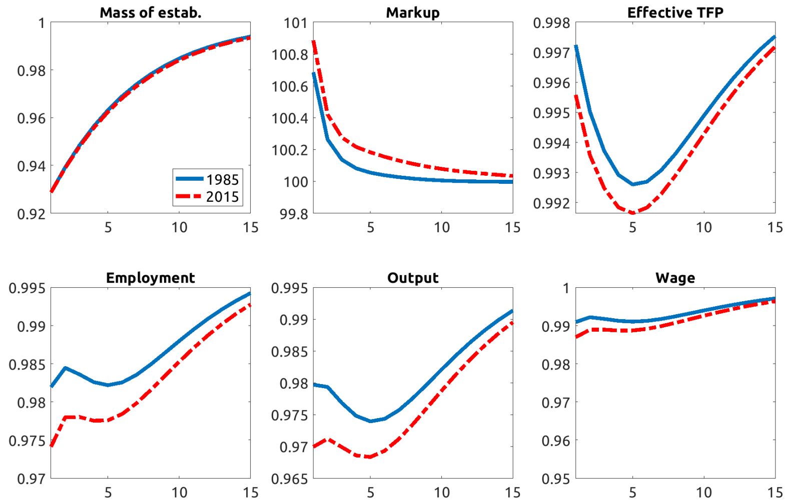
I choose the value of ϵ/σ to match the regression coefficient in 1985 of 0.786 and in 2015 of 0.486. As Table 11 shows, this generates a rise in the wedge between sales and labor reallocation, so that employment growth dispersion as a ratio of sales growth dispersion falls from 60% to 28%. This decline matches the decline of this ratio that I documented in Section 4. 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 is about 20% 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 firms.

How do the effects of an entry shock differ in these two calibrations? Figure 28 depicts the response of each economy to the same transitory, unexpected shock to the mass of potential entrants. As it shows, the markup rises by 75 basis points and only gradually recovers in the 2015 calibration, but in the 1985 calibration, it rises by only 50 basis points and very quickly recovers. Effective TFP falls slightly more in the 2015 calibration. These two effects lead employment to fall in response to the shock by 33% more in the 2015 calibration.

Since the rise in markups following even a temporary fall in entry is long-lasting, this shock combined with a TFP shock has the potential to generate slow employment

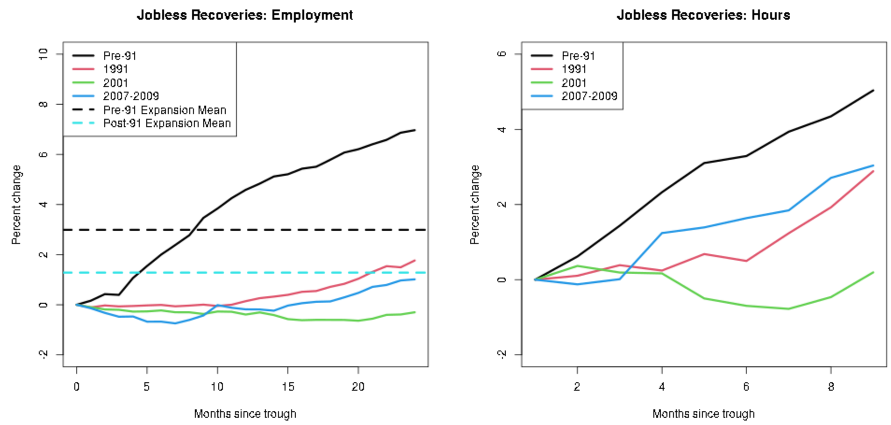
Figure 28: Response to entry shock in 1985 and 2015



recoveries. The trend that I document in the markup–size relationship coincides with the empirically–documented rise in jobless recoveries. Figure 29 depicts the behavior of employment (non-farm payroll) and total hours from the end of NBER recessions. As it shows, employment recovers much more quickly following recessions before 1991 than after. Before 1991, employment recovered slightly more quickly than average following recessions, whereas after 1991, employment stagnated for almost a year before beginning to gradually recover. This timing coincides with the dramatic increase that I document in the markup–revenue relationship. During the Great Recession in particular, while other headwinds may have subsided, the anti-competitive effects of entry continued to buffer the employment recovery.

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 firms' markups become more responsive to their market shares, fluctuations in entry will increase the volatility of aggregate employment.

Figure 29: Jobless Recoveries



8 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 firms. In this paper, I show that incorporating these effects into a general equilibrium, heterogeneous firms model greatly amplifies the effects of entry on aggregate employment and output.

I first present a general equilibrium firm dynamics model with entry and exit, variable elasticity of demand, and adjustment frictions. I calibrate the model to be consistent with the lifecycle of the firm, the adjustment costs of firms, 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 firms 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.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *American Economic Review*, 93(4): 1268–1290.

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.

Crouzet, Nicholas, and Neil R. Mehrotra. 2020. "Small and Large Firms Over the Business Cycle." Mimeo.

Decker, Ryan, John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity : Shocks vs. Responsiveness." Board of Governors of the Federal Reserve System (U.S.) Finance and Economics Discussion Series 2018-007.

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.

Felix, Sonia, and Chiara Maggi. 2019. "What is the Impact of Increased Business Competition?" International Monetary Fund IMF Working Papers 19/276.

Goldberg, Pinelopi Koujianou, and Rebecca Hellerstein. 2013. "A Structural Approach to Identifying the Sources of Local Currency Price Stability." *Review of Economic Studies*, 80(1): 175–210.

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.

Gutiérrez, Germàn, Callum Jones, and Thomas Philippon. 2019. "Entry Costs and the Macroeconomy." *Manuscript.*

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.

Jaravel, Xavier. 2019. "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector." *The Quarterly Journal of Economics*, 134(2): 715–783.

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.

Midrigan, Virgiliu. 2008. "Comment on "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety"." In *NBER Macroeconomics Annual 2007, Volume 22. NBER Chapters*, 355–364. National Bureau of Economic Research, Inc.

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

Moreira, Sara. 2017. "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles." Center for Economic Studies, U.S. Census Bureau Working Papers 17-29.

Nakamura, Emi, and Dawit Zerom. 2010. "Accounting for Incomplete Pass-Through." *Review of Economic Studies*, 77(3): 1192–1230.

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.

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

Suveg, Melinda. 2020. "Did Firm-Exit Affect Prices During the Crisis?" SSRN CEP Discussion Papers.

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

A Compustat Details

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 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 12: 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 13: 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.3 NAICS-2

Fact 1

Table 14: 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

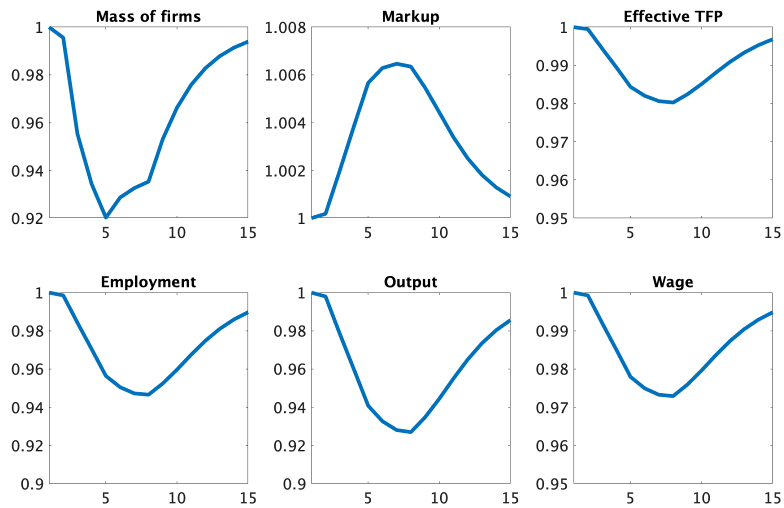
Table 15: Variable input use and relative size over time

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

B 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 30: The Great Recession shock to the entry of firms



C 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 16: Calibrated parameters

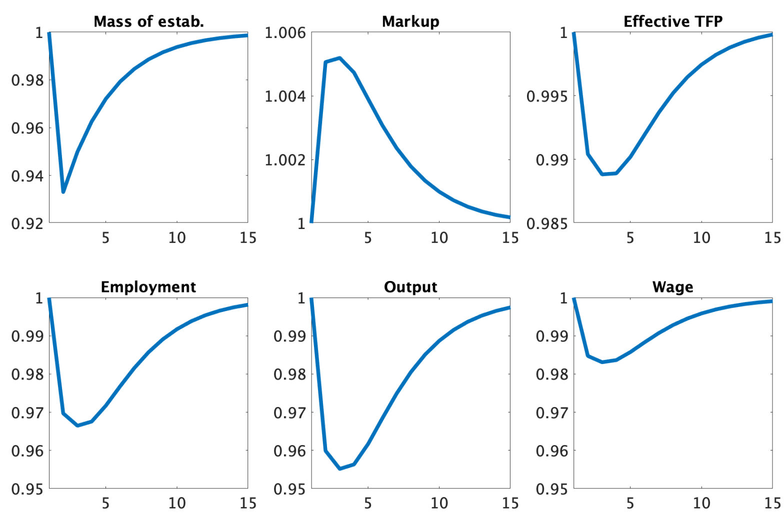
Parameter	Description	Value	Targeted Moment
σ_s	Tfp innovation dispersion	0.29	Labor Dynamism
ϕ_L	Adjustment cost	0.0032	Labor adjustment as fraction of revenue
ϵ/σ	Super-elasticity	0.6	Labor–sales regression
μ_F	Log fixed cost mean	-3.15	Entry rate
σ_F	Log fixed cost dispersion	1.65	Average size exiting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

Table 17: 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%

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

Figure 31: The response of the baseline quantitative model to an MIT shock



D Solution method

D.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.

E 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:

$$G(\tilde{z}; \eta, \theta) = 1 - \left(\frac{\eta}{Az}\right)^\theta \tag{E.1}$$

$$= 1 - \left(\frac{\eta/A}{x}\right)^\theta \tag{E.2}$$

$$= G(\tilde{z}; \eta/A, \theta) \tag{E.3}$$

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 32

This implies that the markup then becomes:

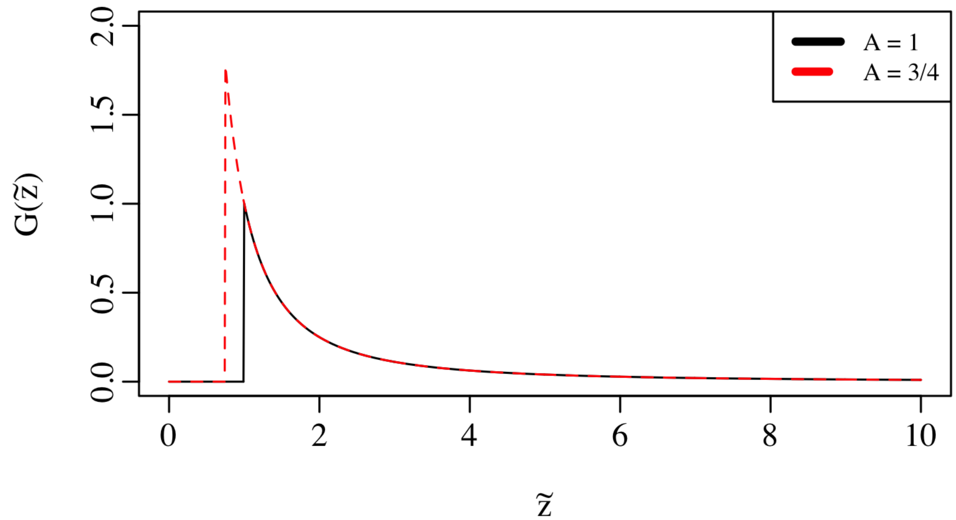
$$\mathcal{M} = \frac{\int_A^\infty \mu(q(\tilde{z})) \frac{q(\tilde{z})}{\tilde{z}} dG(\tilde{z})}{\int_A^\infty \frac{q(\tilde{z})}{\tilde{z}} dG(\tilde{z})}$$

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 32: A change of variables under the Pareto assumption



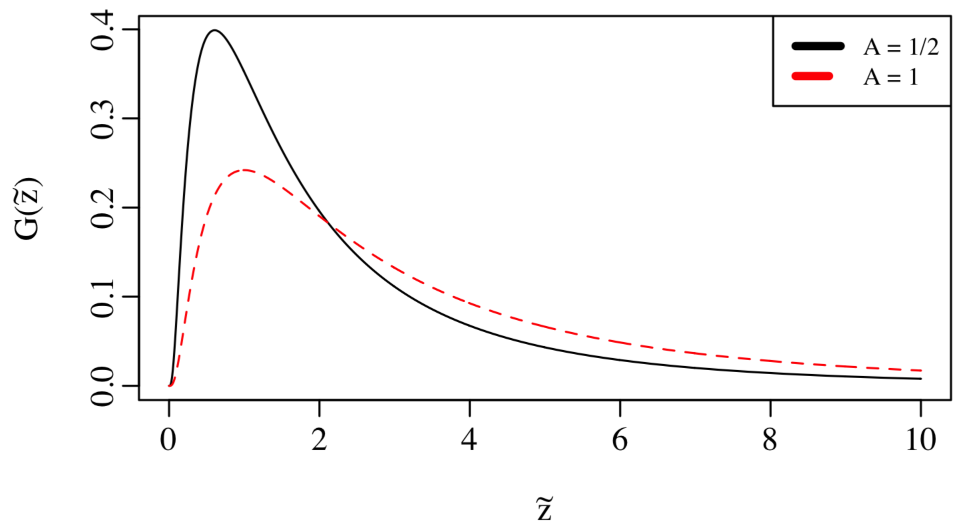
Suppose that $\log z \sim \mathcal{N}(\mu, \sigma^2)$. A change of variables implies that $\log \tilde{z} \equiv \log Az \sim \mathcal{N}(\log A + \mu, \sigma^2)$.

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 33 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 33: A change of variables under the log-normal assumption



F Stochastic Discount Factor

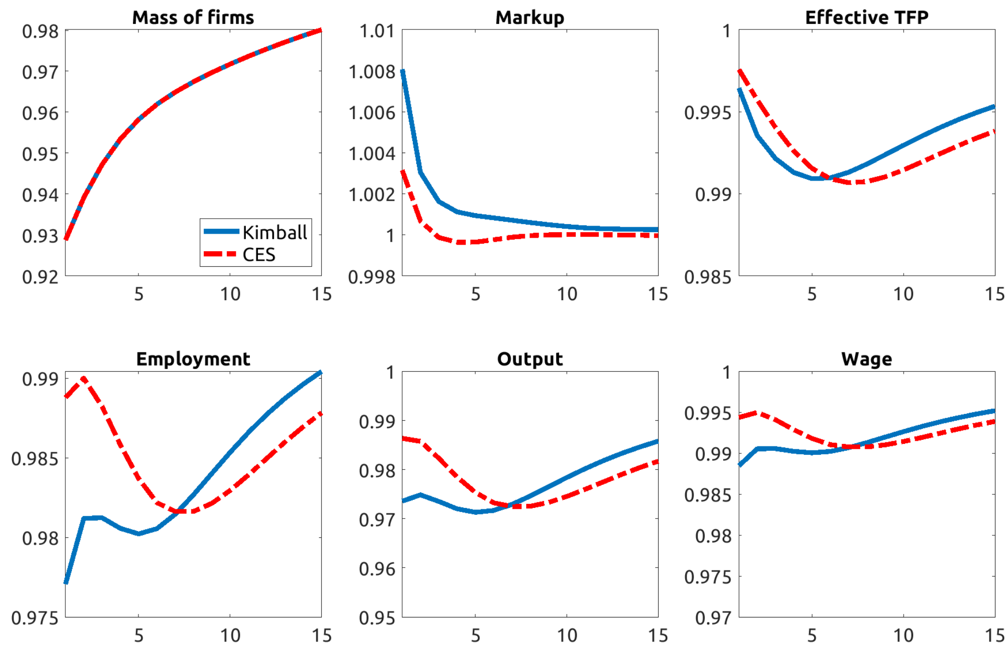
F.1 Shock to entry

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 34. 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 34: Impulse response to an entry shock; variable stochastic discount factor



G 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 18 and 19 summarize the calibration of the free entry model.

I solve for the response of the model economy to a one-time unexpected shock 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 D.1.

In response to the shock, the entry rate and share of employment among entrants and young firms fall. Figure 36 depicts the role of entrants following the shock. The entry rate falls by around 5 percentage points. It recovers quickly, with some over-

Table 18: Calibrated parameters

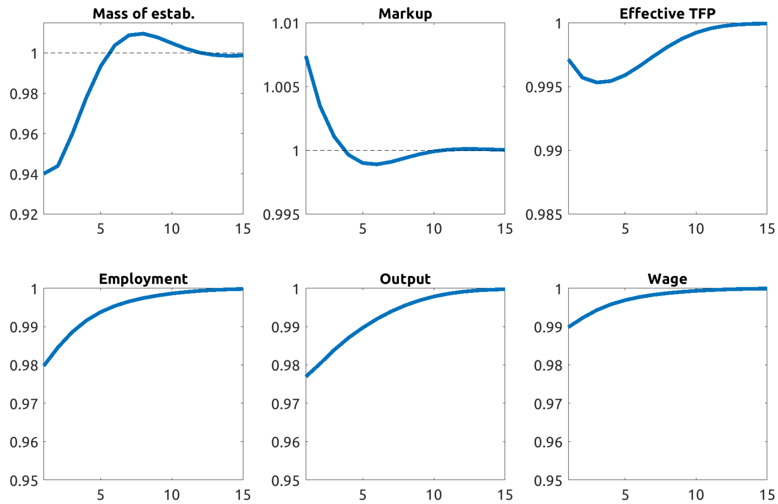
Parameter	Description	Value	Targeted Moment
ρ_s	TFP persistence	0.79	Top 10% share
σ_s	Tfp innovation dispersion	0.17	Var. emp. growth
ϕ_L	Adjustment cost	0.055	Autocorr. emp. growth
ϵ/σ	Super-elasticity	0.57	Labor–sales regression
d_E	Productivity difference of entrants	0.4	Average size entering firm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

Table 19: 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%

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

Figure 35: The response of the baseline quantitative model to an MIT shock
Free Entry Model



shooting, 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 37 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 36: Entrants following the shock
Free Entry Model

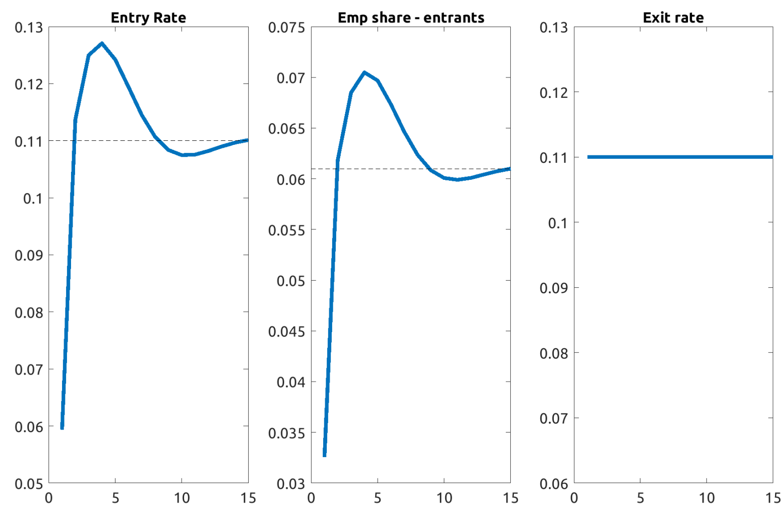


Figure 37: Decomposition of entry shock
Free Entry model