Entry, Variable Markups, and Business Cycles

Click here for most recent version

William Gamber*

December 10, 2020

Abstract

The creation of new businesses ("entry") declines in recessions. In this paper, I study the effects of pro-cyclical entry on aggregate employment in a general equilibrium framework. The key features of the model are that firms' markups increase with market share and adjustment costs prevent employment from reallocating across firms. In response to a decline in entry, incumbent firms' market shares increase and they increase markups and reduce employment. To quantify this mechanism in the model, I study the relationship between variable inputs and revenues 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 35%. I then study shocks to entry in a model that is calibrated to be consistent with this elasticity, 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 3 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 is becoming stronger with time.

*Contact: william.gamber@nyu.edu. I am very grateful to my advisors Simon Gilchrist, Ricardo Lagos, Virgiliu Midrigan for their guidance and support throughout this project. I would also like to thank Jaroslav Borovička, Giuseppe Fiori, Mark Gertler, Sebastian Graves, James Graham, Venky Venkateswaran, and Joshua Weiss for their helpful comments, as well as seminar participants at NYU and the Federal Reserve Board.

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 2014, 5 years after the end of the recession.[1] This fall in entry accompanied a decline in employment of over 6 percent and a slow recovery. In this paper, I study how fluctuations in the creation of new businesses amplify recessionary contractions in employment.

My approach is to study shocks to entry in a general equilibrium firm dynamics model in which market concentration affects markups, aggregate productivity, and employment. I use this model to quantify the following propagation mechanism: a fall in entry causes the market shares of incumbents to rise, leading them to increase markups and reduce employment. I find that, during the Great Recession, this mechanism led the average markup to increase and generated a decline in aggregate employment of 3 percent.

The model I study features heterogeneous firms and endogenous entry and exit decisions. Firms face idiosyncratic, stochastic productivity and a demand curve with an elasticity of demand that declines with relative size, which implies that their markups increase with their market shares. Firms also face adjustment costs that prevent labor from rapidly reallocating across firms and limit firms' responses to idiosyncratic and aggregate shocks.

An important moment in this model is the elasticity of the markup to market share among large firms. To quantify this elasticity, I study the relationship between variable input use and relative sales in a panel dataset on large firms. Under the assumptions that (1) firms can frictionlessly adjust the variable input and (2) markups are fixed, the variable input bill should covary one-for-one with relative sales. I show that the data reject this hypothesis; the typical firm in the sample increases its variable input bill much less than one-for-one with its sales. The structural model incorporates two mechanisms that could generate this pattern: (1) adjustment costs that prevent the firm from fully responding to demand or productivity shocks and (2) a positive relationship between relative sales and markups. I quantify both when I calibrate the model.

I calibrate the parameters underlying the productivity process, demand structure, adjustment costs, and firm lifecycle to match the distribution of firm size and several features of firm dynamics. The model allows me to quantify the size of adjustment costs and the extent to which markups increase with relative sales. In particular, I identify the parameters underlying these mechanisms using the auto-correlation of firm-level

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

employment growth and the regression coefficient of within-firm sales growth on within-firm employment growth. Without adjustment costs, the model implies a counterfactually negative auto-correlation of employment growth, and without a markup-size relationship, the model implies a counterfactually high regression coefficient.

To study the effects of entry on aggregate employment, I then introduce a shock to the mass of potential entrants in the model. This shock is isomorphic to a shock to the ability of entrepreneurs to borrow to finance new firms, and it reduces both the mass of entrants and their average productivity. In the model, a 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, the aggregate markup rises by 0.8% and aggregate productivity falls by 0.5%. Because of these changes, aggregate output falls by 2.5% and employment declines by 2%.

The interaction between adjustment costs and market power is key to generating the rise in the aggregate markup and the contraction in employment. The aggregate markup, defined as the inverse labor share, equals the employment-weighted average of firm level markups. In response to the shock to entry, firms in the model raise their markups, which leads the aggregate 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, reducing the rise in the aggregate markup. Adjustment costs prevent small firms from increasing their employment rapidly and thus inhibit the reallocation mechanism. To quantify the role of adjustment costs, I compare the baseline model to one without adjustment costs, finding that without adjustment costs, reallocation undoes 80% of the immediate rise in the markup.

To study the role of variable markups in this model, I compare the model to one with a constant elasticity of demand. The constant elasticity model implies that markups do not systematically vary with market share. I find that the effects of entry on aggregate employment are doubled in the variable markups economy relative to the constant elasticity model. The difference between the two models arises because falling entry reduces the labor share and leads to a reallocation of output away from high productivity firms only in the variable elasticity model. I conclude that the existing literature understates the importance of firm entry because it ignores the effects of entry on the markups of incumbents.

This paper contributes to a literature on the fall in entry during the Great Recession. That literature observes that, because entrants employ only a small fraction of the total labor force in the US, the decline in entry after 2007 did not significantly deepen the recession (Siemer (2014), Moreira (2017), and Clementi and Palazzo (2016)). I find, however, that the decline in entry during the Great Recession led to a large and persistent contraction in aggregate employment. The reason for this surprising result is that, in my analysis, I account for the response of large incumbent firms to the presence of entrants. In the model I study, large firms increase their markups as their market shares rise. This mechanism implies that, in contrast to previous papers, I find that a contraction in entry has large and immediate effects on aggregate employment and output.

In my model, a shock to entry causes large incumbent firms' market shares to increase, and in response they raise their markups and reduce their employment. This mechanism is economically significant for two reasons. First, large establishments change their markups significantly in response to changes in market share. In general equilibrium, a rise in the markup of a given size generates a larger decline in employment. I find that this mechanism doubles the effects of a shock to entry on aggregate employment relative to models in which markups do not vary with market shares.

My paper also contributes to a recent literature that has found that entry has little or no effect on the aggregate markup in a class of general equilibrium models (Arkolakis et al. (2019) and Edmond, Midrigan and Xu (2018)). In the context of this paper, the key to this neutrality result is a strong reallocation channel: following a decline in entry, all firms increase their markups, but there is also a reallocation of output to low markup firms, which implies that the aggregate markup does not change. These theoretical findings are at odds with causal evidence on the short-run effects of entry on markups and employment (Suveg (2020) and Felix and Maggi (2019)). Moreover, the reallocation channel is inconsistent with the empirical finding that small firms' sales fall by more in recessions large firms' sales (Crouzet and Mehrotra (2020)). In my analysis, I find significant pro-competitive effects of entry, even accounting for firm heterogeneity. The reason for this result is that, in the model I study, adjustment costs prevent the extreme reallocation of output to low-markup firms at business cycle frequencies.

I conclude by turning to two applications of this model. 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 the secular rise in market concentration, 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)). And, even when entrants are the same size as incumbents, introducing heterogeneity into variable markups models reduces and may even completely undo the pro-competitive effects of entry.

A more recent literature argues that entry has little to no effect on the aggregate markup in economies with firm heterogeneity. This result is quite robust. Arkolakis et al. (2019) show that the entire distribution of markups is invariant to changes in trade costs in a class of monopolistic competition models with variable markups, Pareto-distributed productivity, no adjustment costs on inputs, and the existence of a choke price. The reason for this result is that, while domestic producers may lower their markups in response to a fall in trade costs, foreign producers raise theirs. This implies that the univariate distribution of markups is unchanged. Under the Klenow and Willis (2016) preferences I use in this paper, the employment-weighted average markup is unchanged as well. The Bernard et al. (2003) model of Bertrand competition shares a similar feature. In work more similar to this paper, Edmond, Midrigan and Xu (2018) studies an economy that relaxes the choke price assumption and finds that marginal changes in entry have essentially no 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 employment 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.

While this reallocation channel 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' (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, following a long literature in business cycle macroeconomics going back at least to Hopenhayn and Rogerson (1993), I include labor adjustment costs.[2] Labor adjustment costs prevent the extreme reallocation of employment to low markup firms from undoing the 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 surprising 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

[2]As Arkolakis et al. (2019) note, the log-normal assumption is not important.

shows in Swedish data that a one percent increase in exit generated by a restriction in the availability of financing led prices to increase by 1.6 percent. I present a model that is fully consistent with firm heterogeneity, as in Arkolakis et al. (2019) and Edmond, Midrigan and Xu (2018) and these causal estimates of the effect of entry on markups.

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

Another literature studies 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 entry and restrict the size of young firms. A lack of entry and persistence of idiosyncratic conditions generate a "missing cohort" of firms, whose absence from the economy has long lasting effects. Clementi and Palazzo (2016) studies 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 this paper, I build on that literature by incorporating the effects of market concentration. As in the missing cohort literature, in my model entering firms are small relative to incumbents. The innovation in my paper is that in the model, large firms increase their markups in response to the fall in entry. The increase in markups prevents these large incumbent firms from hiring, and so I find that pro–cyclical entry not only lengthens recessions but it also significantly deepens them.

Secular trends in markups and firm dynamism

In this paper, I document a dramatic increase in the strength of the relationship between markups and market share within firms. This fact builds on an existing literature that studies the rise in markups and the fall in the labor share. A central finding of this literature is that the fall in the labor share and rise in markups is driven by a reallocation of output to high markup firms (see, for example Autor et al. (2017), Kehrig and Vincent (2018) and De Loecker and Eeckhout (2017)). A reallocation of output to high markup firms implies a stronger correlation between output and markups.

The main difference between my paper and De Loecker and Eeckhout (2017) is that I study within–firm variation in the markup and allow for greater heterogeneity in production functions. In their approach, it is necessary to assume that firms within an industry share the same production function. My approach allows firms' production functions to vary within industries and over time. I show controlling for heterogeneity across firms increases the measure of how much markups vary with firm size and the extent to which the covariance of markups with firm size has increased over time.

Kehrig and Vincent (2018) study firm- and establishment-level data on the labor share in manufacturing. They document that the decline in the aggregate labor share was driven by a reallocation to output to low-labor share firms. They show that this reallocation was driven by a stronger negative relationship between firm size and their labor shares. These findings are consistent with the facts I document about falling within-firm relationship between variable input demand and revenue. I account for this relationship in my model with a rise in the relationship between demand elasticity and firm size.

In contrast to recent papers studying entry and markups, I highlight the role of firm dynamics at business cycle frequencies. This emphasis might seem at odds with a view that the rise in markups was driven by the rise in "superstar" firms that are permanently large. It is consistent, however, with Kehrig and Vincent (2018), who document that high markups are quite transitory; 60% of high markup businesses in an average year are no longer high markup five years later.

Much of the existing theoretical literature on the secular trends in markups studies its causes and welfare consequences. There are many papers in this literature, but some include Edmond, Midrigan and Xu (2018), Baqaee and Farhi (2020), and Weiss (2020) who study the welfare costs of the rise in markups and Gutierrez and Philippon (2018) who study the effects of rising markups on investment. A smaller literature links the rise in markups to business cycles. Wang and Werning (2020) and Mongey (2017) study how market structure affects monetary non-neutrality. My paper is the first to study how the changing relationship between firm size and markups affects how pro-cyclical entry propagates to aggregate outcomes.

My paper also contributes to a literature studying the decline in labor dynamism. Decker et al. (2018) document declining labor dynamism and argue that it is consistent with rising adjustment frictions. In this paper, I connect the rise in the markup-size relationship to declining labor reallocation. This offers a new explanation for this trend and naturally suggests that the rise in concentration should affect employment dynamics over the business cycle. I show in the last section of this paper that the model can account for both the rising markup–size relationship and the decline in labor reallocation.

2 Background: 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 1977–2014.

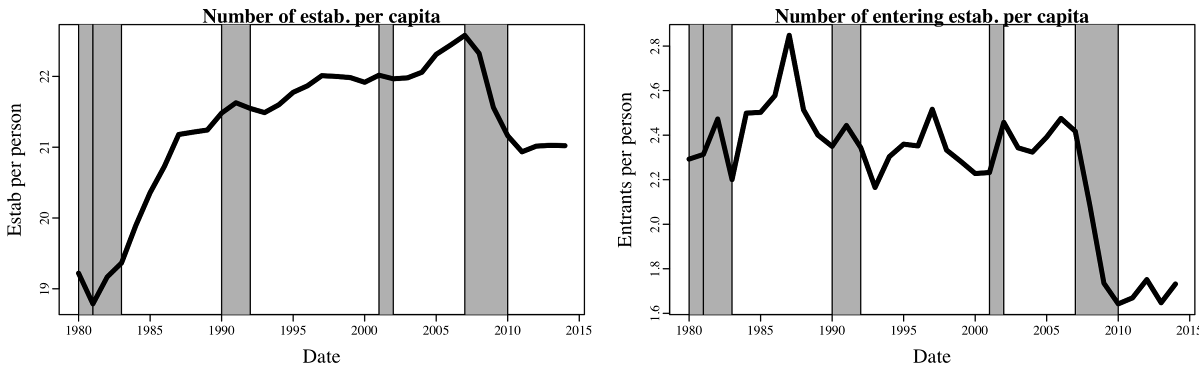
Entry rates in the typical recession

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. The left panel of Figure 1 shows the number of entering establishments, and the right panel shows the annual log growth rate of the number of establishments each year in the BDS. The creation of new establishments varies pro-cyclically over the sample. Net entry (the growth rate in the number of firms) 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 firms, and the fall in the number of firms during the Great Recession was especially 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.[3] 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. Both are pro-cyclical. Since pro-cyclical exit rates imply counter–cyclical net entry, the fall in the number of firms during recessions is driven by the entry margin rather than by rising exit.

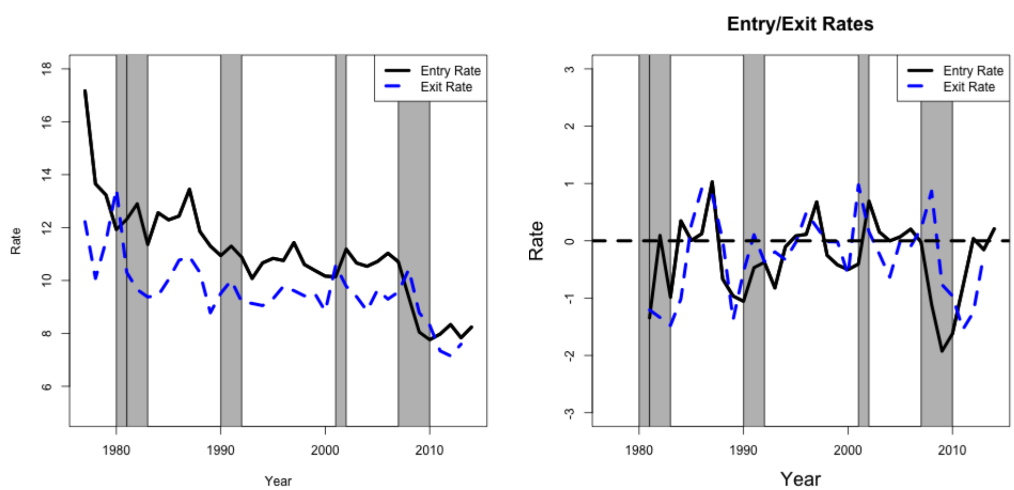
[3]Note that the BDS does not directly report the number of exiting firms. Instead, I infer the number of exiting firms by noting that the change in the number of firms in a given year must equal the number of entering firms less the number of exiting firms.

Figure 1: Growth in the number of establishments in the BDS



NBER Recessions shaded in grey

Figure 2: Entry and exit of establishments in BDS



NBER Recessions shaded in grey

Table 1: Entrants relative to the whole economy, 1985–2014

Moment	Firms	Establishments
Entry rate	10.3%	10.4%
Emp. share entrants	2.9%	5.4%
Emp. share young	23%	40%
Relative size of entrants	28%	51.9%

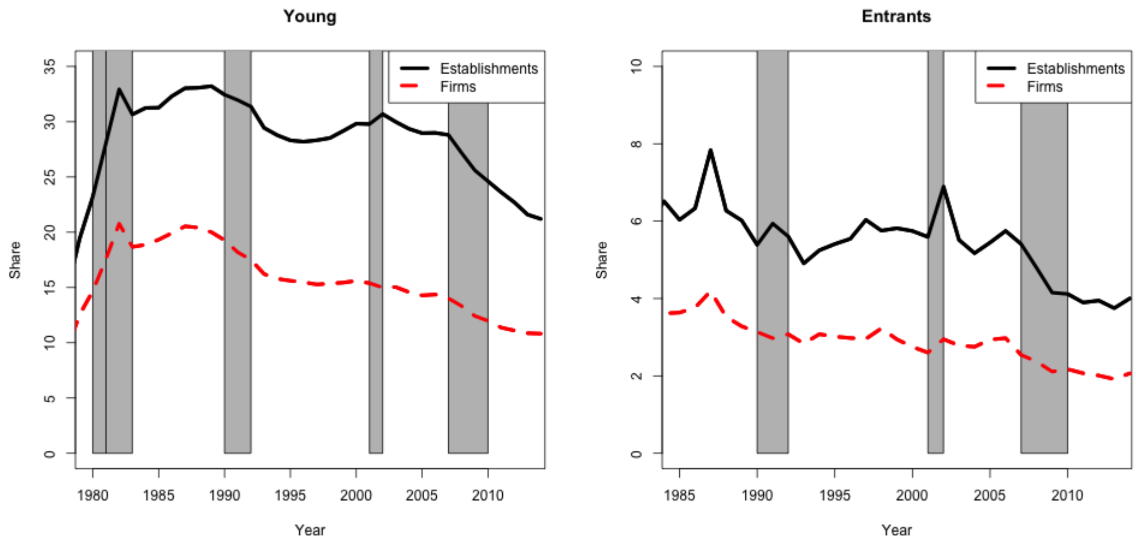
Given that these are aggregate fluctuations, they mask considerable heterogeneity in business dynamism across industries. They are, for example, muted relative to the fluctuations in manufacturing plants documented by Lee and Mukoyama (2015), who find that entry rates are 4.7% lower in recessions than they are in booms. They also find that exit rates are 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 employment shares of young and entering firms are pro–cyclical over the sample depicted, 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: Employment share of young and entering businesses



3 Markups and market share among large firms

The goal of this paper is to quantify the following mechanism: the fall in the entry rate during recessions leads incumbent firms' market shares to rise, and so they increase their markups and reduce their employment. The effect of missing entry on incumbent market shares is pinned down by the relative size of entrants and the distribution of firm size. So, in this section, I quantify the second part of the mechanism. I provide direct evidence that large firms increase their markups as their market shares rise. I also show theoretically how markup variation relates to employment. I will use the estimates of the size of this relationship to calibrate the quantitative model I study later.

Empirical framework

Consider a firm with a production function in a variable input L and a static input K.[4] The distinction between variable and static inputs is that the 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 demand or productivity, which I summarize with A. The production function can be expressed as:

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

$$Y = Q(A; K, L) \tag{3.1}$$

Denote by α the output elasticity of the variable input L. This coefficient might vary over time or across firms and industries. For simplicity of exposition, I proceed using a Cobb–Douglas production function. However, the results all hold in a more general class of production functions[5].

$$Y = AK^{\theta}L^{\alpha} \tag{3.2}$$

The firm can frictionlessly hire any amount of the flexible input at price W. The dual problem of the firm is to minimize the cost of producing a given level Y of output:

$$\min_{V,K} WL + rK + F \tag{3.3}$$

$$\text{such that } Y \leqslant Q(A; K, V) \tag{3.4}$$

Denote by λ the Lagrange multiplier on the output constraint. This is equal to the marginal cost, since it is the value of relaxing that constraint. A first order condition of the cost minimization problem with respect to V is then

$$W = A\alpha\lambda K^{\theta}L^{\alpha-1} \tag{3.5}$$

Multiplying both sides by L gives

$$WL = \alpha\lambda Y \tag{3.6}$$

The markup μ equals the price of the output good P divided by the marginal cost, λ. Substituting into the first order condition then gives a relationship between total variable input cost, revenue, the markup, and the output elasticity.

$$WL = \alpha\frac{PY}{\mu} \tag{3.7}$$

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:

$$\log WL = \log\alpha + \log PY - \log\mu \tag{3.8}$$

Consider the following regression:

[5] See De Loecker and Eeckhout (2017) for a more complete discussion of this approach.

$$\log WL = \tilde{\alpha} + \beta \log PY + \epsilon \tag{3.9}$$

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

$$\beta = 1 - \frac{\text{Cov}(\log PY, \log \mu)}{\text{Var}(\log PY)} \tag{3.10}$$

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 deviation of this coefficient from 1 is informative about the degree to which 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. I exclude financial firms and utilities, and for my baseline results I classify firms using the Fama-French-49 industry definition.[6]

The sample of firms, while not representative of the average firm in the economy, covers a large portion of US output and employment. Compustat firms 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 variable costs, 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 the cost of goods sold (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:

$$\log(WL)_{ift} = \alpha_{g(ift)} + \beta \log(PY)_{ift} + \epsilon_{ift} \tag{3.11}$$

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 captures how, at a business cycle frequency, firms' variable input use varies when their revenues change. I find values well below 1 for these regressions, varying between 0.356 for employment and 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, Structural Interpretation 1

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.

Structural 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) account for firm heterogeneity and show that markups increase by more if we instead use within–firm variation. 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. Column (1) understates the extent to which firms increase their markups as they grow because it misattributes variation in markups to permanent variation in output elasticities across firms.

Markups vs. Production Function

In interpreting these regression coefficients, I allow for a variety of specifications 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 necessarily 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

$$\frac{K_{ift}}{L_{ift}} = \alpha_{it} + \beta P_{ift} Y_{ift} + \epsilon_{ift} \tag{3.12}$$

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

To avoid misattributing variation in this wedge entirely to variation in the markup, I allow for labor adjustment costs in the structural model I study later. In a simulated method of moments exercise, I jointly estimate both a structural parameter that 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.

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 firms or over time. This allows me to estimate how markups vary with revenue and 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 any fixed effect. 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) imply.

The rise in the markup-revenue relationship

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

Figure 4 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 monotonically decline by significant amounts from 1985 to 2015. The plots using XLR exhibit noisier estimates[8] but still generally decline after 2000. 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

[8] This is not surprising given the sparsity of data available for that measure.

Figure 4: Variable Input–Revenue Relationship, Rolling Windows

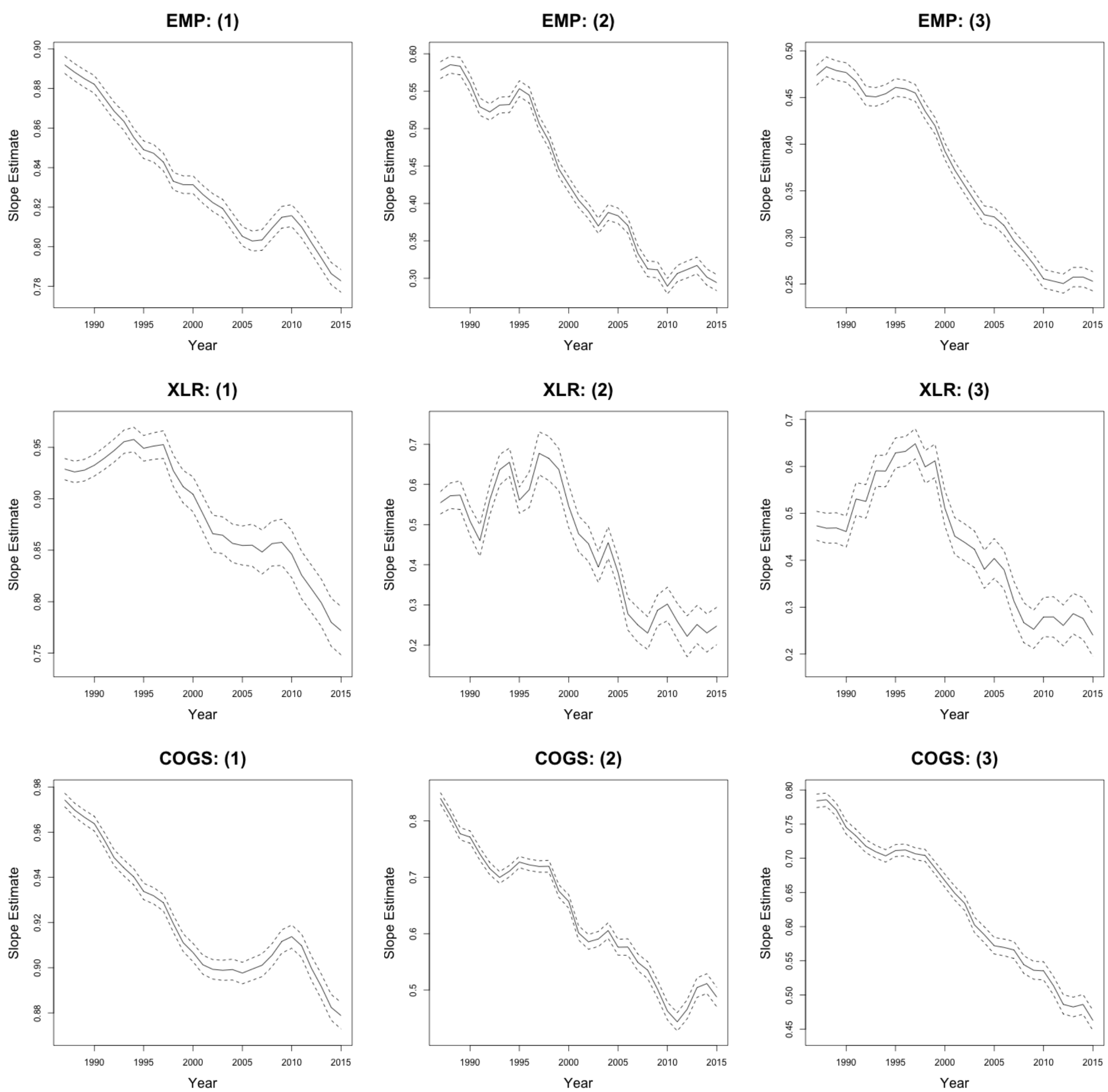


Table 5: Variable input use and relative size over time

	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

Table 6: Dynamism in Compustat, 2010

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

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.

Markups and labor reallocation

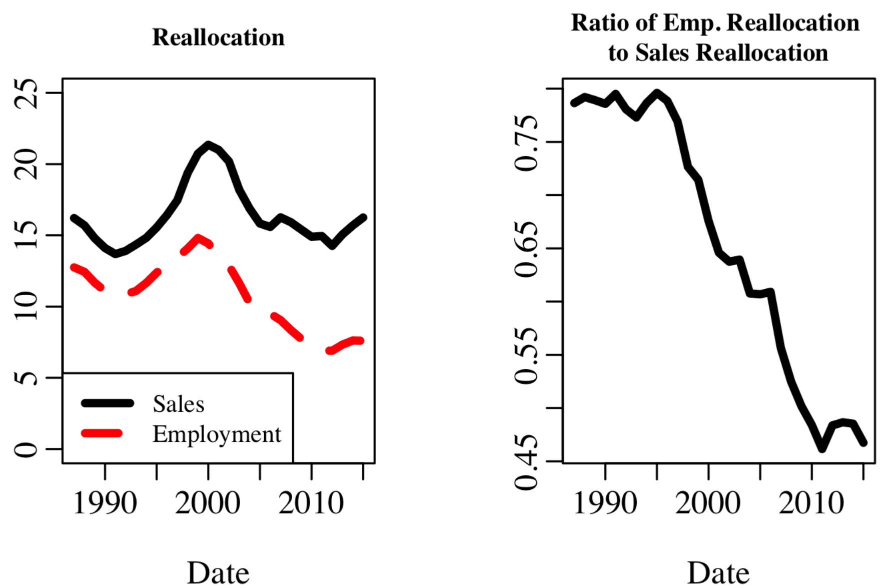
In this section, I provide suggestive evidence that the rise in the relationship between markups and revenue explains the fall in labor reallocation documented by Decker et al. (2018). Differencing the first order condition of the firm over time gives a decomposition of the cross–sectional variance of sales growth ("sales reallocation") into the variance of employment growth ("employment reallocation") and two terms about markup variation:

$$\underbrace{\mathrm{Var}(\Delta \log PY)}_{\text{Sales reallocation}} = \underbrace{\mathrm{Var}(\Delta \log WL)}_{\text{Employment reallocation}} + \underbrace{\mathrm{Var}(\Delta \log \mu) + 2\mathrm{Cov}(\Delta \log \mu, \Delta \log L)}_{\text{Markup variation}} \tag{3.13}$$

Inspecting this decomposition shows that there is a tight relationship between the cross-sectional dispersion in labor and sales growth, mediated by markup dispersion. A positive markup-size relationship and variation in the markup within firms both imply a wedge between these two measures, and so the higher correlation between markups and firm size that I document could drive a rise in the wedge between employment reallocation and 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

Figure 5: Employment and Sales Dynamism



and its covariance with employment growth.

These measures have not been stable over time. As emphasized in Decker et al. (2018), employment reallocation has fallen after a surge in the mid 1990s. The red line in Figure 5 confirms this decline in Compustat. A less–studied fact is that sales reallocation has remained stable over that period, and 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 80% of sales reallocation, it has fallen to 45%.

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 show in a general equilibrium model that the rising covariance between markups and firm size can quantitatively explain this rising wedge.

Summary

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.

In a static framework, 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.

4 Quantitative Model

In this section, I develop a general equilibrium firm dynamics model to study business cycle fluctuations in entry. In the model, heterogeneous firms' markups vary with their market shares. The model generates this behavior through a demand system that features an elasticity of demand that falls with relative output. The firms in this model face labor adjustment costs that prevent the reallocation of labor following idiosyncratic and aggregate shocks. The model also features endogenous entry and exit decisions. As in the data, entering firms start smaller than incumbents on average and grow over time. Entry and the number of competing firms affects markups, the labor share, aggregate employment, and aggregate productivity.

Environment

Time is discrete and continues forever. There are three types of agents in this economy: a representative household who consumes a final good and supplies labor, a final goods producer who uses a continuum of intermediate inputs to produce the final good, and 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 present discounted value of future utility:

$$\sum_{t=0}^{\infty} \beta^t u(C_t, L_t) \tag{4.1}$$

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. The household period budget constraint is thus:

$$C_t \leqslant W_t L_t + \Pi_t \tag{4.2}$$

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

$$W_t = -\frac{u_{L,t}}{u_{C,t}} \tag{4.3}$$

Final goods producer

A perfectly competitive representative firm produces the final consumption good using as inputs a continuum of measure N_t of intermediate goods, each indexed by ω. The final goods producer takes as given the prices of the intermediate goods and minimizes the cost of producing output. Their production technology will imply an elasticity of demand that increases with the relative quantity they choose of each differentiated input. This production function takes the following form:

$$\int_0^{N_t} \Upsilon\left(\frac{y_t(\omega)}{Y_t}\right) d\omega = 1 \tag{4.4}$$

where $\Upsilon(q)$ is a function that satisfies three conditions: it is increasing $\Upsilon'(q) > 0$, concave $\Upsilon''(q) < 0$ and is 1 at the point 1: $\Upsilon(1) = 1$. Given quantities of each intermediate variety $\{y_t(\omega)\}$, the production function implicitly defines the quantity of output Y_t. For the main exercises in this paper, I use the Klenow and Willis (2016) specification of $\Upsilon(q)$:

$$\Upsilon(q) = 1 + (\sigma - 1)\exp\left(\frac{1}{\epsilon}\right)\epsilon^{\frac{\sigma}{\epsilon}-1}\left[\Gamma\left(\frac{\sigma}{\epsilon}, \frac{1}{\epsilon}\right) - \Gamma\left(\frac{\sigma}{\epsilon}, \frac{q^{\epsilon/\sigma}}{\epsilon}\right)\right] \tag{4.5}$$

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

$$\Gamma(s, x) = \int_x^{\infty} t^{s-1}\epsilon^{-t} dt \tag{4.6}$$

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.

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

$$p_t(\omega) = \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right) D_t \tag{4.7}$$

where D_t is termed the "demand index" and is defined as

$$D_t \equiv \left(\int_0^{N_t} \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right)\frac{y_t(\omega)}{Y_t} d\omega\right)^{-1} \tag{4.8}$$

Under the Klenow and Willis (2016) specification, the demand curve is then (up to a scaling constant):

$$\Upsilon'(q) = \frac{\sigma - 1}{\sigma} \exp\left(\frac{1 - q^{\frac{\epsilon}{\sigma}}}{\epsilon}\right) \tag{4.9}$$

The elasticity of demand under this assumption is equal to $\sigma q^{-\frac{\epsilon}{\sigma}}$. The demand elasticity declines with the quantity chosen of the intermediate good, and the elasticity of the elasticity of demand to quantity produced is the ratio ϵ/σ, often denoted the "superelasticity."

Intermediate goods producers

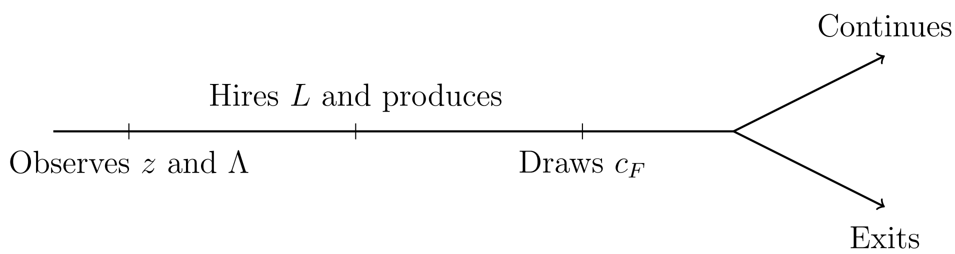
Each period, a variable mass N_t of intermediate goods producers ("establishments") each uses labor to produce a differentiated good. Each establishment is the sole producer of its differentiated variety ω. They each have access to a production function in labor L and idiosyncratic productivity z:

$$F(L; z) = zL^{\eta} \tag{4.10}$$

Establishments must pay a random i.i.d. fixed cost $\phi_F \sim G_F$ to operate each period. If a firm chooses not to pay the random fixed cost, it exits. The value of exit is normalized to 0. Firms are also exogenously destroyed at rate $\gamma > 0$. Finally, firms face labor adjustment costs $\phi(L, L')$.

The information structure and timing are summarized in Figure 6. A firm enters period t having employed L_{t-1} workers in the previous period. It observes its idiosyncratic productivity z_t and then chooses L_t. It receives period profits π_t and pays adjustment cost $\phi(L_t, L_{t-1})$. After producing, it then draws a fixed cost of production and decides whether to immediately exit or to pay the cost and continue producing in the next period. With probability γ, the fixed cost it draws is infinite and the firm will choose exit with probability 1.

Figure 6: Timing for incumbent establishments



Let Λ summarize aggregate states that are relevant to each establishment. The recursive problem of an incumbent establishment who employed L employees last period, has productivity z and has paid fixed cost ϕ_F is listed below. It discounts future streams of profits using the discount factor m.[9]

$$V(L,z;\Lambda)=\max_{p,L'}\pi(z,L',p;\Lambda)-c(L',L)+\int\max\left\{0,\tilde{V}(L',z,c_F;\Lambda)\right\}dJ(c_F) \tag{4.11}$$

$$\tilde{V}(L,z,c_F;\Lambda)=-c_F+\beta(1-\gamma)\mathbb{E}\Big[m'V(L,z';\Lambda)\big|z\Big] \tag{4.12}$$

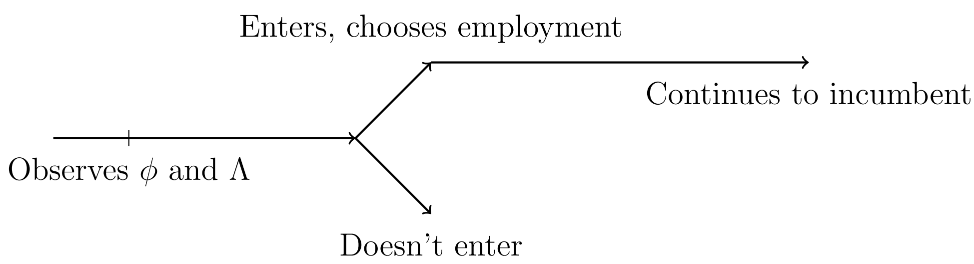
$$\pi(z,L',p;\Lambda)=\left(p-\frac{W}{L}\right)d(p;\Lambda) \tag{4.13}$$

$$y\leqslant zL \tag{4.14}$$

[9]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 H, 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 7: Timing for potential entrants



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 7 summarizes the information structure for potential entrants.

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

$$V_E(\phi) = \int_z \max_L \beta(1-\phi)\mathbb{E}\Big[V(z,L)|\phi\Big]dH(z|\phi) \tag{4.15}$$

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 $\hat{\phi}$.

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 $\Lambda(z, \ell)$.

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 and the policy functions of the firms.

Aggregation

In spite of the heterogeneity present in this model, it aggregates to a representative firm economy.[10] Consider the aggregate production function, where Z_t denotes *aggregate productivity*:

$$Y_t = Z_t L_t \tag{4.16}$$

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

$$Z_t = \left(\int \int \frac{q_t(z,L)}{z} d\Lambda_t(z,L) \right)^{-1} \tag{4.17}$$

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:

$$\mathcal{M}_t = \frac{Y_t}{W_t L_t} \tag{4.18}$$

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

$$\mathcal{M}_t = \int \int \mu_t(z,L) \frac{\ell_t(z,L)}{L_t} d\Lambda_t(z,L) \tag{4.19}$$

[10] Though, solving the model still requires approximating the value function of the firms. See Appendix D.2 for details.

5 Steady state

In the steady state of this model, firms are heterogeneous along a number of dimensions. Each firm's value is defined by its two states, 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.

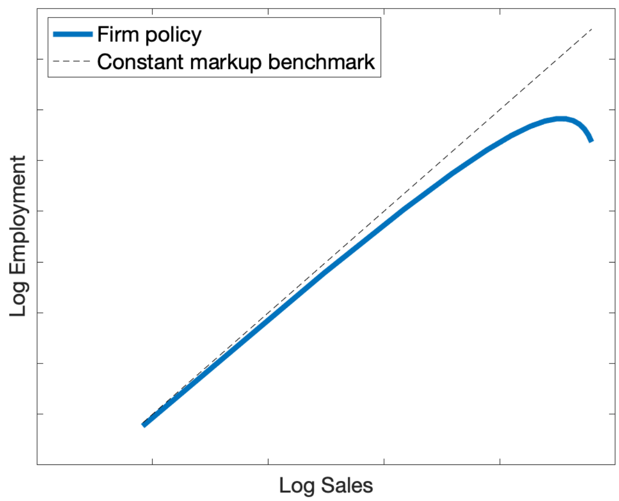
I calibrate the model to the behavior of establishments. Establishments are more likely to represent unique products and thus might be better thought of as the relevant unit of competition for this model. As Table 1 shows, entering establishments are larger relative to incumbent establishments than entering firms are relative to incumbent firms. This means that entering and young establishments employ a larger fraction of workers than do entering and young firms. In Appendix B, I explore an alternative calibration of the model in which the unit of production is a firm rather than an establishment.

The employment-sales regression

As I showed in Section 3, large firms change their variable input use less than one-for-one with revenue, which suggests that their markups increase with their market share. In the model, two forces generate this pattern: (1) the superelasticity of demand means that large firms have more market power to set markups over marginal cost, and (2) adjustment costs prevent firms from adjusting their variable inputs in response to productivity shock. The size of the adjustment cost is disciplined by data on labor adjustment variation, and so I choose the superelasticity to target the regression coefficients from Section 3.

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 the firm's productivity rises, it produces more and its sales rise. In response to the rise in sales, it increases its markups. The increase in markups means that the firm increases its employment less than one-for-one with its sales. Figure 8 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 Kimball model increase their markups as their sales grow, which implies that the slope of the policy function is always less than one. The policy function is also concave. This arises because larger firms increase their markups more with sales than do small firms. Eventually, establishments become so productive that when they experience positive productivity shocks, they produce more, increase their prices and

Figure 8: Employment and sales in the frictionless model



revenues, and employ fewer people.

The relationship between employment and revenue is not linear in the model: the employment–sales regression coefficient is smaller for large firms than it is for small firms. This presents a challenge in calibrating the model, since the average Compustat firm is larger than the average firm in the economy. To calibrate the model, I ensure that the regression coefficient estimated using Equation (3.9) on a sample of the largest firms in the model equals that in the data.

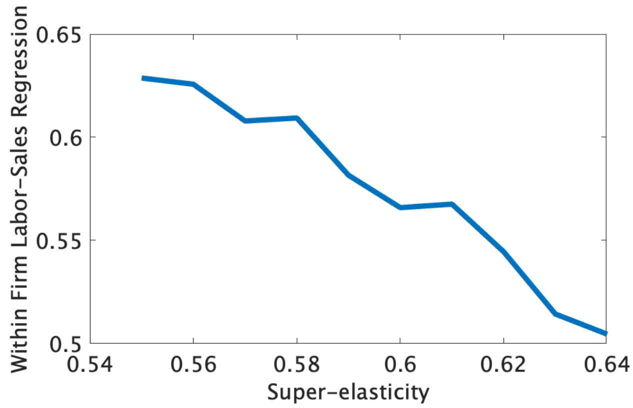
The sample I use in Compustat covers about 1% of firms and 30% of U.S. non–farm payroll. In my simulated method of moments estimation procedure, I simulate a sample of firms in the model and then estimate the regression on a subsample of the top 1% of firms by sales in the model economy. This procedure generates a comparable subsample to estimate the super-elasticity. In Figure 9, 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:

Figure 9: Identification of the superelasticity



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

These 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:

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

I assume that productivity follows an AR(1) process in logs, with persistence ρ_z and innovation variance σ_z^2. The signal that potential entrants receive about their future productivity is Pareto distributed. Figure 10 depicts the distributions of the signal and of realized productivity. To ensure that large entrants are not driving the results, I truncate the Pareto distribution. The productivity realization conditional on the signal follows the same AR(1) law of motion that productivity follows:

$$z = \rho_z q + \sigma_z \epsilon \quad \epsilon \overset{iid}{\sim} \mathcal{N}(0,1) \tag{5.3}$$

Calibration strategy

I exogenously set some parameters and then jointly choose seven of them to target important moments. The pre-set parameter choices are summarized in Table 7. I then estimate the remaining parameters using simulated method of moments, jointly choosing productivity innovation persistence and dispersion ρ_z and σ_z, the adjustment

Figure 10: The distribution of the signal and productivity

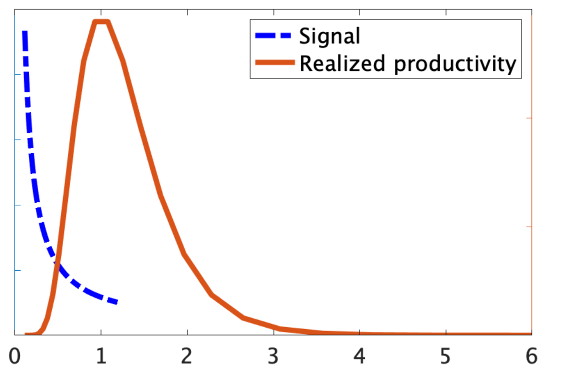


Table 7: Pre-set parameters

Parameter	Description	Value	Source/Target
β	Discount factor	0.96	Annual model
$\mathbb{P}(\text{exit})$	Probability of exit	0.11	Annual entry rate
σ	Kimball demand elasticity	10	
γ	Exogenous exit rate	1.5%	
M	Mass of entrants	1	Normalization
ν	Inverse Frisch Elasticity	0.5	
δ	Job separation rate	0.19	

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 equal $\exp(0)$ and the fixed cost of production equal to 0 with probability $(1 - \mathbb{P}(\text{exit}))$ and infinity with probability $\mathbb{P}(\text{exit})$. In the appendix, I discuss a calibration with a non-degenerate distribution of fixed costs that features selection on exit.

While these parameters affects several moments in the model, they each intuitively correspond to one or two particular moments. The persistence of productivity and dispersion in its innovation affect the cross–sectional variance of firm–level log sales growth, which I estimate to be 15% in Compustat, and the share of sales among the 10% largest firms. The Pareto parameter for the entrant signal affects the relative size of entering firms.

The calibration strategy allows me to jointly estimate the adjustment cost and the super-elasticity of demand and thus to distinguish between two mechanisms that could account for the Compustat reduced-form regression results. I identify the degree of

Table 8: Calibrated parameters

Parameter	Description	Value	Targeted Moment
ρ_s	TFP persistence	0.79	Top 10% share
σ_s	Tfp innovation dispersion	0.29	Var. emp. growth
ϕ_L	Adjustment cost	0.07	Autocorr. emp. growth
ϵ/σ	Super-elasticity	0.6	Labor–sales regression
ξ	Signal Pareto tail	0.95	Average size entering firm
σ	Elasticity parameter	20	Average markup

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 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 0.55, which matches the coefficient using specification (3) and COGS in a 5-year window centered around 2005. Table 16 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 17 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 roughly in line with that in the data. The model also fits the share of employment at entrant and young establishments that I estimate in the BDS. Fitting these 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 estimation strategy for the super-elasticity of demand is novel in that it relies on within–firm variation in sales and markups among a sample of large firms. Still, 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

Table 9: Calibration Targets & Model Fit

Moment	Target	Source	Model moment
Var($\Delta \log L$))	6.17%	Compustat	5.82%
Var($\Delta \log PY$))	14.15%	Compustat	13.4%
$\rho(\Delta \log L_t, \Delta \log L_{t-1})$	0.13	Compustat	0.14
$\rho(\Delta \log P_t Y_t, \Delta \log P_{t-1} Y_{t-1})$	0.12	Compustat	0.12
Labor–sales regression	0.654	Compustat	0.628
Average size of entering firm	50%	CP	0.526%
Top 10% share of sales	75%	Compustat, industry average	69%
Frac. rel. sales. below 1	79%	Compustat, industry average	80%
Cost–weighted average markup	1.25	DLE	1.2645
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

and Konings (2019), Berger and Vavra (2019), and Gopinath, Itskhoki and Rigobon (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, increasing the elasticity of markups to revenue from 0.07 to 0.35. Setting the superelasticity to its value in Edmond, Midrigan and Xu (2018) of $\epsilon/\sigma = 0.14$ in my model implies a model markup elasticity of 0.12, closer to the Compustat regression that does not allow for heterogeneity across firms within an industry.

Consistent with these "micro" estimates, my estimated value of $\epsilon/\sigma = 0.6$ 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 and a negative relationship between employment and revenue among large firms.

Table 10: Selected parameterizations of the superelasticity

Paper	ϵ/σ	Note
This paper	0.6	
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

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 superelasticity 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 in the new model, and estimate the regression coefficient, I find $\hat{\beta}_L = 0.652$. The labor adjustment cost reduces the extent to which firms increase their employment with their size, but it only accounts for 6.5% of the deviation of the coefficient from 1.[11]

Aggregate parameters

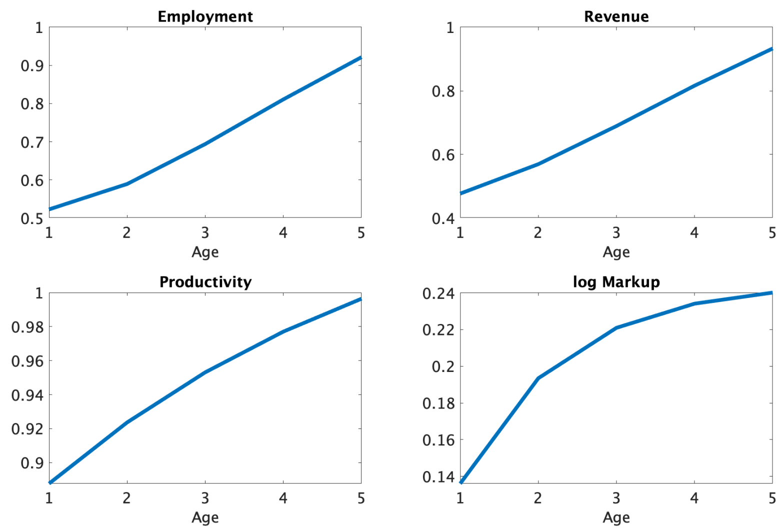
There are some parameters whose values do not affect the steady state of the economy, only its response to aggregate shocks. These are the inverse Frisch elasticity, which I set to be $\nu = 1/2$ and the disutility of labor parameter, ψ, which I set so that the steady state wage is 1.

The lifecycle of the firm

Firms in the model, as in the data, begin small and grow slowly. Figure 11 shows that employment and revenue at entering establishments are around 50% of the the average

[11]Setting the labor adjustment cost to 0 very strongly affects the autocorrelation of employment growth. With the labor adjustment cost set to 0, this moment is -0.215. This is the essence of the identification strategy.

Figure 11: The lifecycle of the firm in the quantitative model



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 mean reverts slowly 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 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.

Markups and concentration

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 markup in the model is 1.27, which is far below its value of 1.65 in the data. The cost-weighted markup is the relevant measure of the distortions due to markup, which is why I choose to target that value in the calibration.

Figure 12 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.

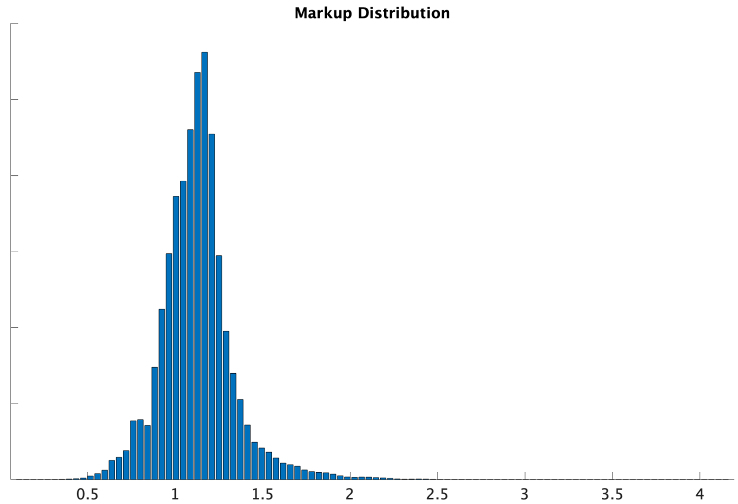


Figure 12: The cost-weighted distribution of markups

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.[12] In the main exercise, the shock lasts for one year and has no persistence. After the initial shock is realized, the households and firms 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.2.

[12]Why not a shock to the cost of entry? The selection mechanism present in this model means that a rise in the entry cost produces a counterfactual rise in the average productivity of entrants. This means that the share of employment at entrants and young firms does not fall very much. I discuss this further in Appendix F.

Entry shocks as financial shocks

As Siemer (2014) documents, financial conditions affect the creation of new businesses. I do not take a stance on the specific origin of the shock in the model, but it is consistent with a shock to the ability of entering firms to obtain financing. Suppose that entering firms must borrow to cover the sunk cost of entry. Moreover, suppose that the entry signal is private to the potential entrant and that it cannot credibly convey its value to the lender. Then, a reduction in credit will reduce the mass of potential entrants equally across values of their productivity signal.

In the appendix, I explore alternative shocks. In Appendix F, I study a shock to the cost of entry. In Appendix G, I introduce a financial friction and study a shock to the cost of issuing equity, following Gilchrist et al. (2017).

An entry shock

Figure 13 depicts the response of the baseline quantitative model to a shock to the mass of potential entrants. The shock only lasts for one year, but its effects are persistent because it takes the economy time to build back up the mass of establishments.

The fall in entry leads the mass of establishments to decline and the market shares of incumbents to rise. In response, incumbents increase their markups, the cost–weighted average markup rises by 75 basis points. Since the labor share is the inverse of the average markup, it falls by 75 basis points. Effective TFP, the ratio of output to aggregate employment, falls endogenously by nearly 50 basis points. Employment falls by 2 percent, and output falls by around 3 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 falls. Figure 14 depicts the role of entrants following the shock. The entry rate falls by around 3 percentage points. The fall in the entry rate is typical for a recession, as noted by Lee and Mukoyama (2015). 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 5.5% to just below 3%. Because young firms are more likely to have low productivity and are thus more likely to exit, the contraction in entry leads to a slight decline in the average entry rate.

The share of employment at young firms falls by around 10 percentage points, in line with the data on the Great Recession. Even though the employment share at entering firms rebounds quickly, the employment share at young firms remains persistently below its steady state level.

Figure 13: The response of the baseline quantitative model to an MIT shock

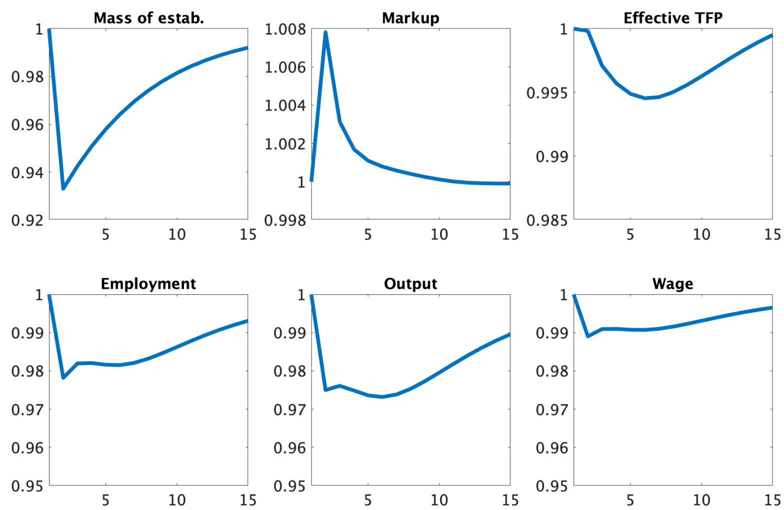
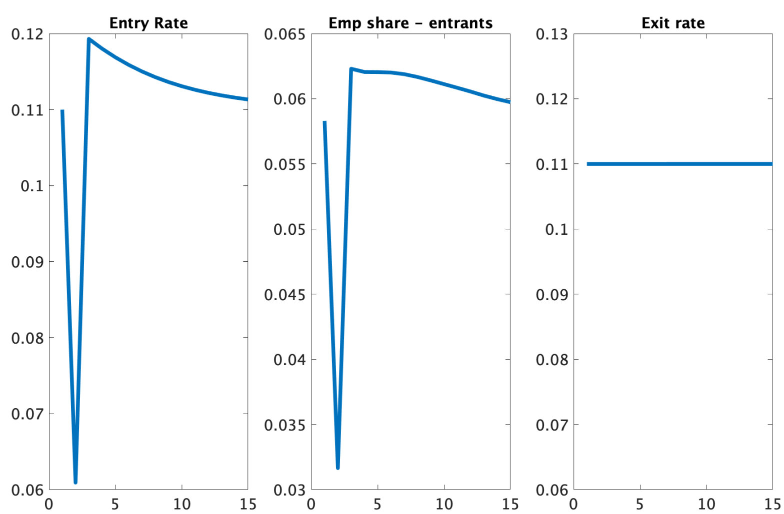


Figure 14: Entrants following the shock



Markups and Productivity

To understand the role 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:

$$
\begin{aligned}
Y_t &= Z_t L_t \\
\mu_t &= \frac{Y_t}{W_t L_t} \\
W_t &= \psi L_t^{\nu}
\end{aligned}
\tag{6.1}
$$

Given paths for the cost–weighted markup μ_t and aggregate effective productivity A_t, equations (6.1) imply paths for output Y_t, employment L_t, and the wage W_t. Note that altering the paths for μ_t or A_t and recomputing these aggregate quantities does not represent an equilibrium of this economy. It does, however, allow me to decompose the equilibrium paths of the aggregate variables.

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 the aggregate equations above can be expressed as labor supply and labor demand equations:

$$
\begin{aligned}
\log W_t &= \log \psi + \nu \log L_t \\
\log W_t &= \log A_t - \log \mu_t
\end{aligned}
\tag{6.2}
$$

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 15 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.

The rise in the markup of 75 basis points generates a 1.5 percent decline in employment, with the remainder of the decline in L_t due to the fall in effective TFP. Figure 16 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 most of the immediate decline in employment, but the markup returns to its steady state level more quickly than TFP. TFP accounts for nearly all of the decline in employment after 5 years.

Figure 15: A rise in the markup or a fall in effective TFP

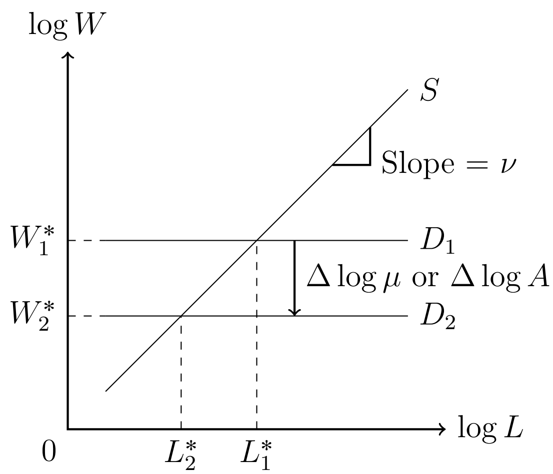


Figure 16: Decomposition of entry shock

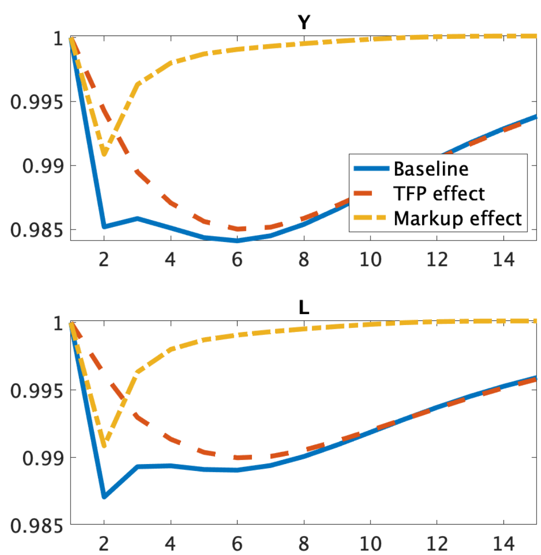
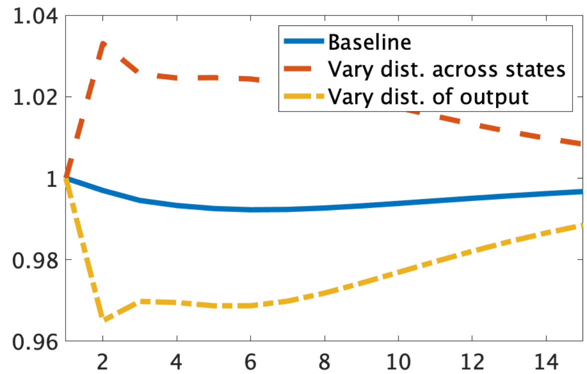


Figure 17: A TFP decomposition



Aggregate TFP

The decline in aggregate TFP accounts for a large portion of the contraction in employment. 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:

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

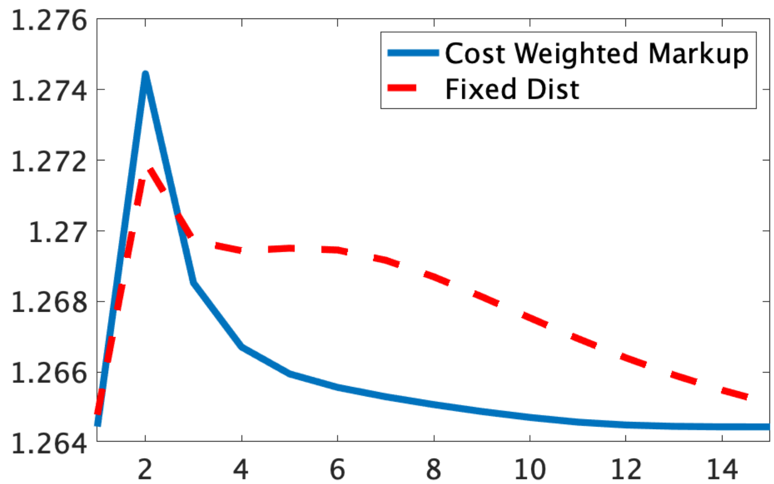
In an accounting sense, Z_t might fluctuate because of changes in $q_t(z,L)$ or changes in $d\Lambda_t(z,L)$. Figure 17 decomposes the path of TFP into each of these two changes. The red dashed line holds fixed the function $q_t(z,L) = q_{SS}(z,L)$ but allows the distribution $d\Lambda_t(z,L)$ to vary. TFP in this exercise rises because entrants are less productive than incumbents, and so a fall in entry leaves the economy with fewer unproductive establishments.

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.

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:

Figure 18: Decomposition of the response of markups



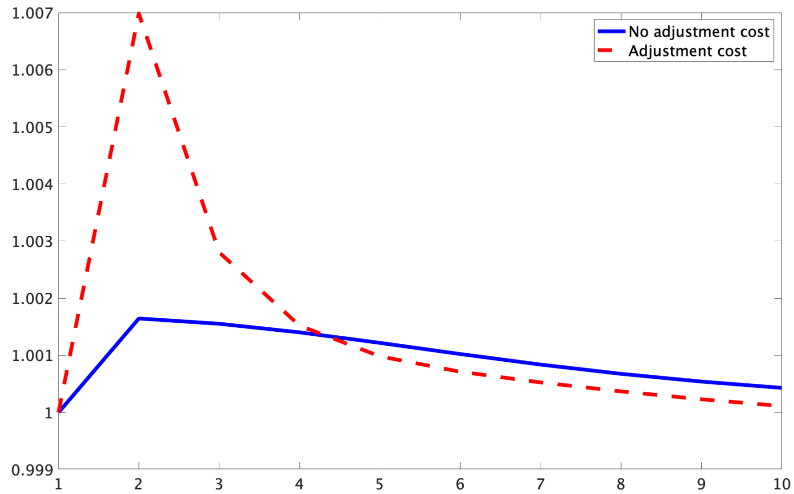
$$\mathcal{M}_t = \int\int \mu_t(z,L)\frac{\ell_t(z,L)}{L_t}d\Lambda_t(z,L) \tag{6.3}$$

The shock to entry primarily affects the markups of individual firms $\mu_t(z)$, but it also affects 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 prevent the reallocation channel from immediately reducing the aggregate markup. Figure 18 depicts the results of a decomposition of the path of the cost weighted markup. In red, I allow markups to vary and hold costs and the distribution fixed. This shows that the average firm raises its markups in response to the shock. The blue line shows the path of actual markups, relative to their steady state. There is reallocation to small, low-markup, firms following the shock because they face a higher elasticity of demand, which implies that they benefit more from the fall in entry.

Because this is not a nominal model, I do not study inflation. However, the large rise in markups at the firm level suggests that the fall in entry could help explain the missing deflation during the last recession. Countercyclical markups lead inflation to rise. Any inflation measure that does not fully account for substitution effects would measure a rise in inflation that is larger than the cost–weighted markup increase.

Figure 19: The role of the adjustment cost in reallocation



The role of adjustment costs

Adjustment costs act to prevent some of the reallocation of output to low-markup firms. To quantify this mechanism, I solve for the impulse response to the same shock in an economy without adjustment costs. In Figure 19, I plot the difference between the unweighted and weighted markup movements as a fraction of the unweighted markup movement in each economy. As it shows, without adjustment costs, reallocation undoes 60% of the increase in the markup, and the adjustment costs in the baseline model prevent around a third of that reallocation effect.

The interaction of adjustment costs and market power in this paper is a novel mechanism. Typically, the effects of entry on markups in Kimball models are undone by a reallocation towards small firms, as in Edmond, Midrigan and Xu (2018). However, in this model, adjustment costs imply that small firms are not willing to hire, and so output is not reallocated as strongly to those firms. In summary, adjustment costs prevent small firms from hiring, while the increase in market power dissuades large firms from hiring.

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 of their theorem in a few ways: productivity is not Pareto distributed, there are adjustment costs, and there is no choke price. Adjustment costs, as I discussed, inhibits most of the reallocation effect. The distributional assumption turns out to take care of the rest.

To see why, first observe that entry has almost no effect on the cost–weighted markup in Edmond, Midrigan and Xu (2018) either. The Pareto distribution plus the fact firms face no adjustment costs implies that a change in entry affects the distribution of markups in a very particular way. A fall in entry effectively scales the underlying Pareto distribution of productivity. Because of the properties of the Pareto distribution, the scaled distribution is the same Pareto distribution, with a higher lower bound. In this model, the smallest firms do not produce much, and so 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 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

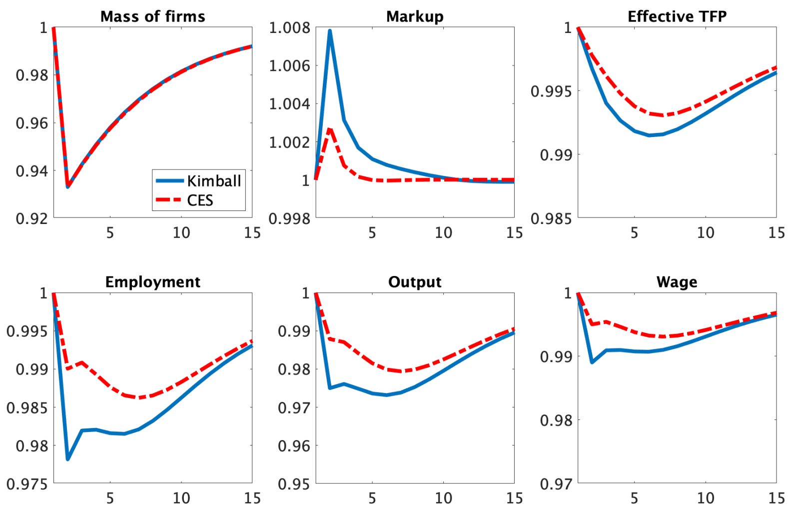
To quantify the role of variable markups in the propagation of entry shocks to aggregate employment and output, I compare the Kimball model to one in which establishments' demand elasticities do not vary with their market shares. This comparison model features constant elasticity of substitution (CES) 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

$$\Upsilon(q) = q^{\frac{\sigma-1}{\sigma}} \tag{6.4}$$

I subject each economy to the same entry shock as before. Figure 20 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 25 basis points) in the CES model because adjustment costs push firms away from their frictionless optimal solution. However,

Figure 20: Entry shock in the Kimball and CES models

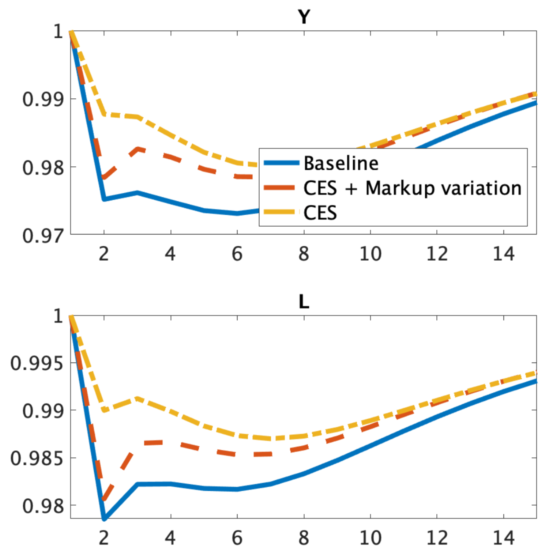


the rise in the markup is small compared to the rise in the Kimball model, meaning that employment in the CES model does not fall as sharply as in the Kimball model. Because of the love-of-variety effects present in both models, effective TFP falls by a similar amount in each. However, in the Kimball 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.

Decomposition of the role of variable markups

To decompose the differences between the Kimball and CES model into the part arising from the aggregate markup and the part arising from aggregate TFP, I study the paths of aggregate variables in the variable–markup economy, feeding in the path of the aggregate markup or aggregate TFP from the constant–markup economy. Figure 21 depicts the results of this decomposition. As it shows, about half of the difference between these two economies is due to the difference in the path of TFP and the other half to the markup.

Figure 21: Decomposition of the difference between Kimball and CES economies



7 Quantitative applications

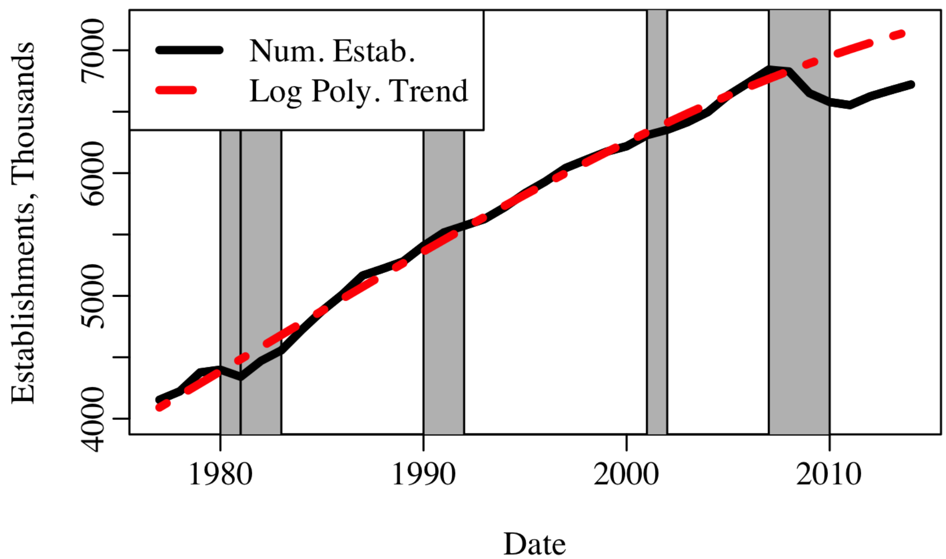
In this section, I study two applications of the model. In the first, I study the role of entry in the sharp contraction in employment during the Great Recession and its subsequent slow recovery. I show that an entry shock that reproduces the path of the mass of firms during the Great Recession leads 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 3. 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 22 shows the path of the number of establishments since 1977 and a log polynomial trend estimated on data before the Great Recession. As it shows, during the Great Recession, the number of establishments fell by 4 percent relative to 2007 and fell by nearly 6 percent relative to its trend. While the number of firms typically

Figure 22: Number of firms/establishments relative to 2007



falls in a recession, these declines were 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 above 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 24 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 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

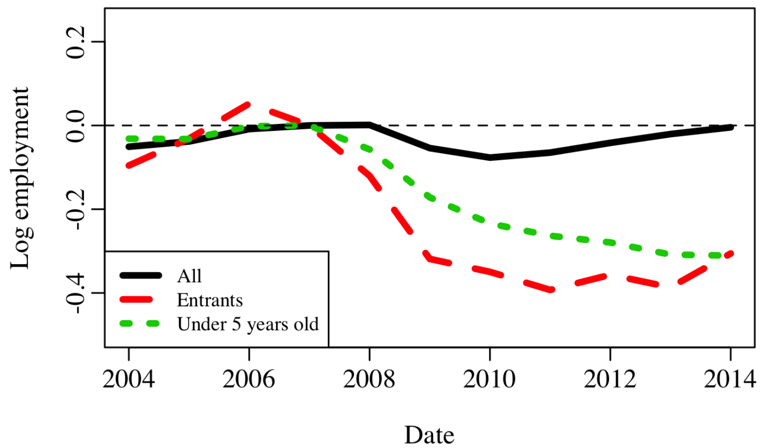


Figure 23: Age and employment among establishments

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 underlying sequence of shocks to the mass of potential entrants implies a decline of 30 percent in the first 2 years in the mass of potential entrants and a gradual recovery thereafter. To compare these shocks to a measurable counterpart, I use data from the U.S. Census Bureau's Business Formation Statistics (BFS) database. The BFS classies business applications as "high propensity" if they are likely to become businesses with a payroll[13]. As Figure 25 shows, the decline is slightly more than the fall in "high propensity" business applications. I also show the decline in the number entering establishments in the BDS, which falls by 35 percent over that period.

The rising importance of markups for business cycles

As I showed in Section 3, 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.

[13]Includes applications (a) from a corporate entity, (b) that indicate they are hiring employees, purchasing a business or changing organizational type, (c) that provide a first wages-paid date (planned wages); or (d) that have a NAICS industry code in manufacturing (31-33), retail stores (44), health care (62), or restaurants/food service (72). See https://www.census.gov/programs-surveys/bfs/technical-documentation/glossary.html for details.

Figure 24: The Great Recession

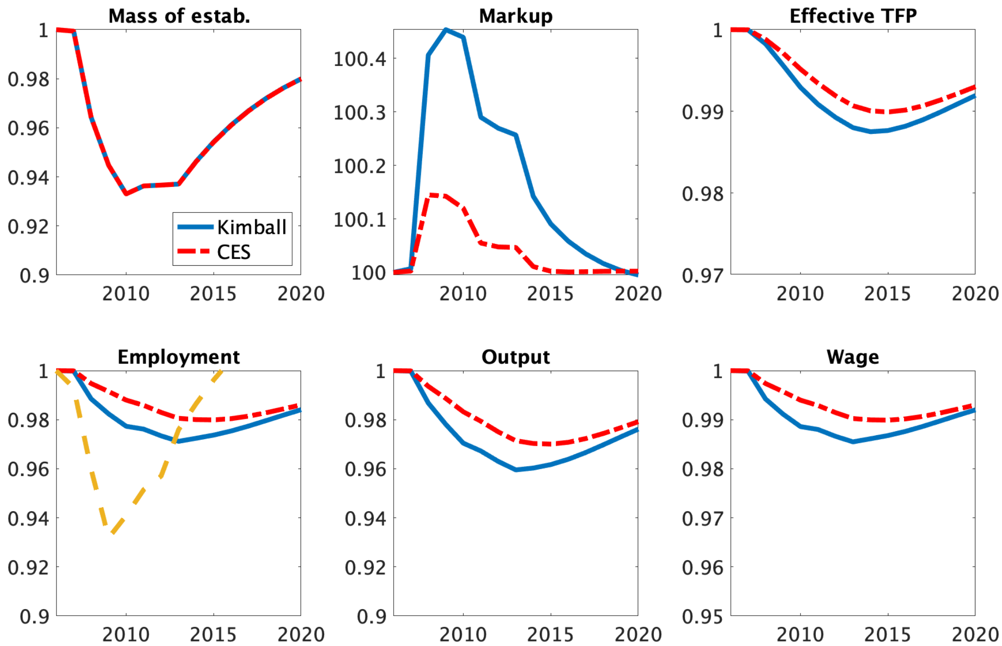


Figure 25: The mass of potential entrants

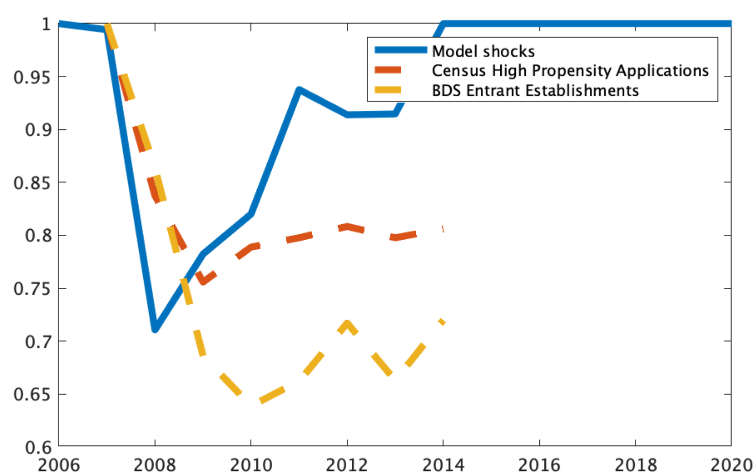


Table 11: Selected moments, 1985 vs 2015 calibration

Calibration	ϵ/σ	β_L	Labor rea./Sales rea.	Cost-weighted markup
1985	0.45	0.763	60%	1.254
2015	0.75	0.466	28%	1.273

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 3. 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 26 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 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 recoveries. The trend that I document in the markup–size relationship coincides with the empirically–documented rise in jobless recoveries. Figure 27 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

Figure 26: Response to entry shock in 1985 and 2015

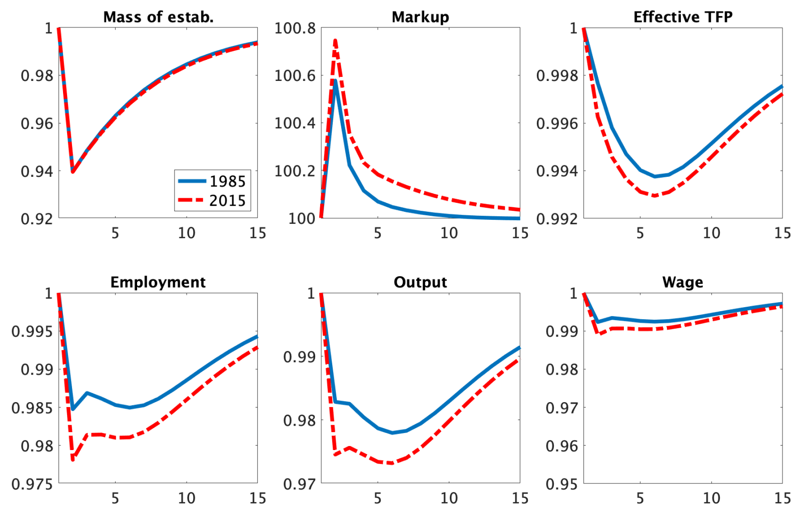
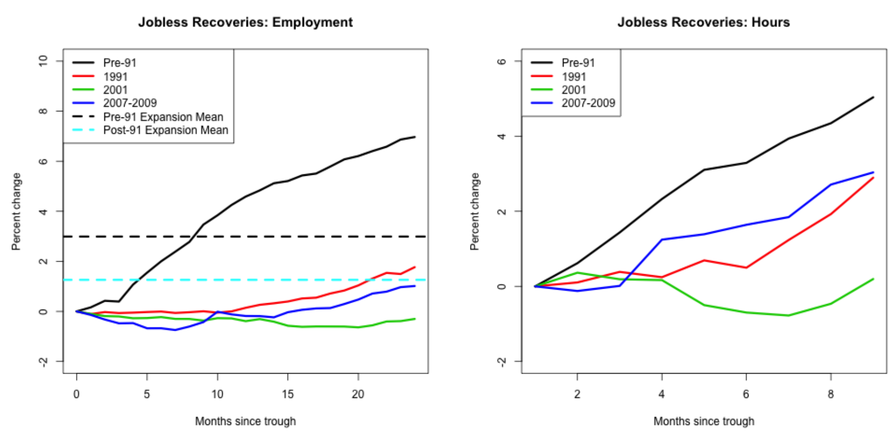


Figure 27: Jobless Recoveries



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.

8 Additional evidence on entry and markups

The theory that I propose finds that entry has significant effects on the aggregate markup. In this section, I present non-causal evidence that is consistent with this mechanism. I combine information on the revenue share of variable inputs at the 2-digit sector level from Compustat with 2-digit entry rates from the universe of U.S. firms and establishments in the BDS.

Entry and variable input shares

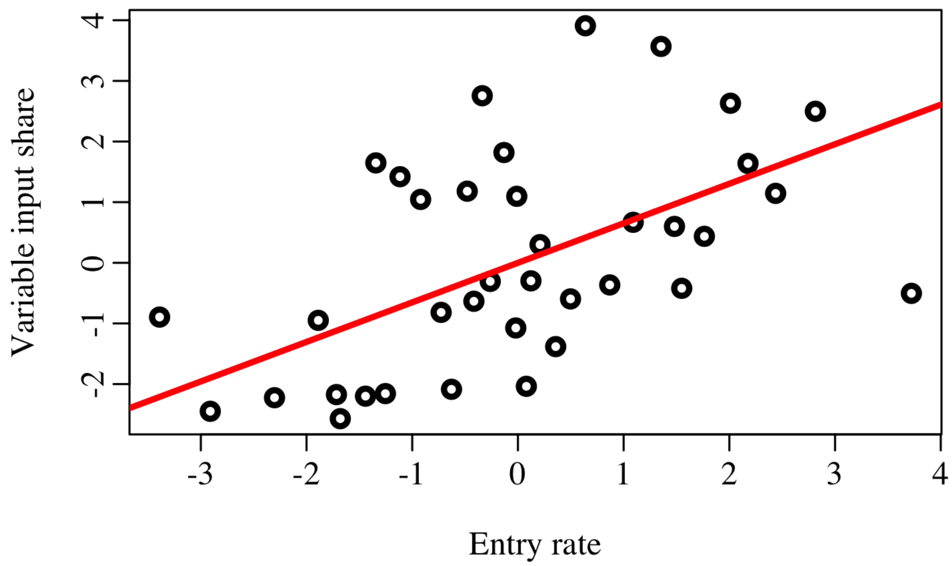
I first estimate

$$\frac{\sum_{f\in i} P^V_{i,f,t} V_{i,f,t}}{\sum_{f\in i} P_{i,f,t} Y_{i,f,t}} = \alpha_i + \beta \times (\text{Entry Rate})_{i,t} + \epsilon_{i,t} \tag{8.1}$$

I estimate this regression using COGS as the measure of variable input cost and using the establishment-level entry rate from the BDS. I winsorise the data, dropping observations with extreme values for either variable.[14] The estimated coefficient is 0.652 (0.005***): a 1 percentage point increase in the entry rate relative to the average within an industry is associated with a 42 basis point rise in the variable input share of revenue. Figure 28 shows the associated binscatter plot. These numbers are quantitatively in line with the findings from the theory.

[14] Though, results are robust to not winsorising.

Figure 28: Binscatter plot of the entry rate and variable input share



9 Conclusion

Competitive conditions change dramatically in recessions. These changes were especially large during the Great Recession, when the number of operating firms fell by 6 percent and of operating establishments fell by 4 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 evidence that large firms increase their markups significantly as their revenues grow. I find large estimates of the elasticity of markups to revenue, ranging from 21.7% to 64.4%. My preferred specification implies that, among large firms, a firm whose revenues grow at double the rate of its competitors within its industry increases its markups by around 35%. This finding suggests that fluctuations in the market power of incumbents could be quantitatively important for business cycles.

I then study entry and business cycles in a model that is consistent with this estimate. The model rationalizes the markup-revenue relationship with a demand system in which elasticities fall with relative output. I calibrate the model to be consistent with the lifecycle of the firm, the adjustment costs of firms, and labor reallocation, as well as the regressions I estimate in the panel data. I find that a fall in entry generates large falls in employment and output. The fall is nearly doubled relative to a model with constant markups, which cannot account for the markup-size relationship I document. The difference between these two models is due to the rise in the cost-weighted markup and the reallocation away from high-productivity producers in the variable markups model.

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, any model of countercyclical markups implies that inflation may not fall much in recessions. Because of the reallocation toward low markup firms in this model, this model implies that firms raise their markups by more than the aggregate markup increases. 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.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and

John Van Reenen. 2017. “The Fall of the Labor Share and the Rise of Superstar Firms.” Centre for Economic Performance, LSE CEP Discussion Papers dp1482.

Baqaee, David Rezza, and Emmanuel Farhi. 2020. “Productivity and Misallocation in General Equilibrium.” *The Quarterly Journal of Economics*, 135(1): 105–163.

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 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.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. 2017. "Inflation Dynamics during the Financial Crisis." *American Economic Review*, 107(3): 785–823.

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.

Gutierrez, German, and Thomas Philippon. 2018. "Ownership, Concentration, and Investment." *AEA Papers and Proceedings*, 108: 432–437.

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

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.

Kehrig, Matthias, and Nicolas Vincent. 2018. "The Micro-Level Anatomy of the Labor Share Decline." National Bureau of Economic Research, Inc NBER Working Papers 25275.

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.

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.

Mongey, Simon. 2017. "Market Structure and Monetary Non-neutrality." Federal Reserve Bank of Minneapolis Staff Report 558.

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.

Wang, Olivier, and Iván Werning. 2020. "Dynamic Oligopoly and Price Stickiness." National Bureau of Economic Research, Inc NBER Working Papers 27536.

Weiss, Joshua. 2020. "Intangible Investment and Market Concentration." *Working Paper*.

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

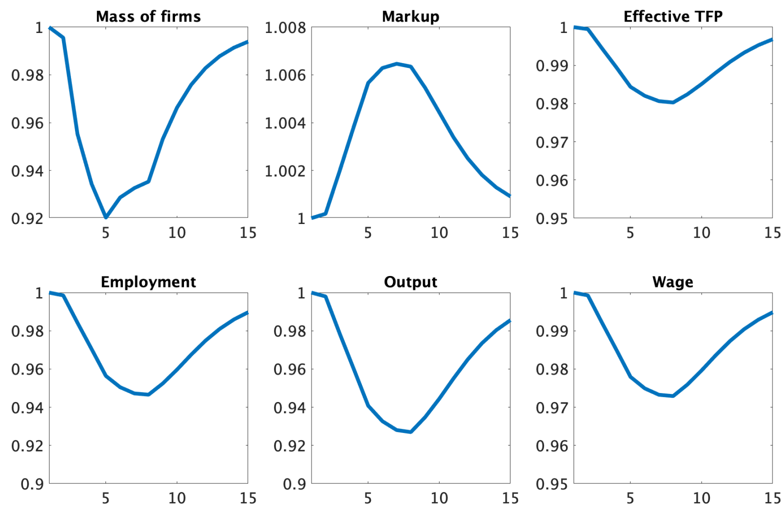
Table 15: Variable input use and relative size over time

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 29: 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

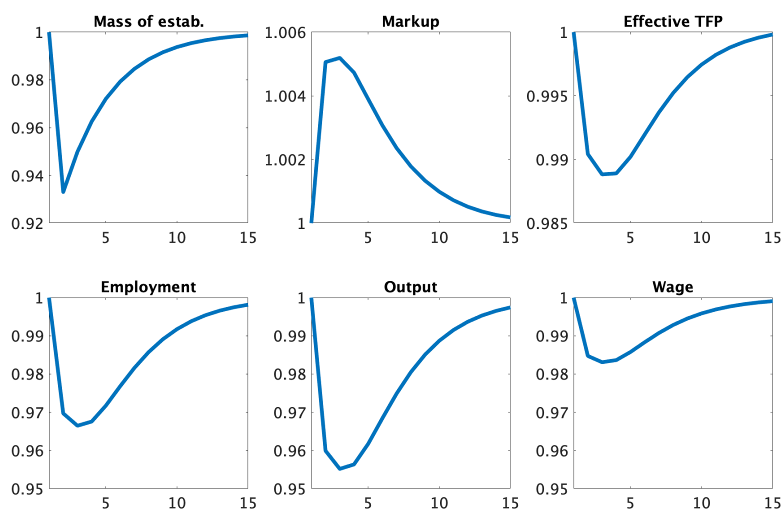
Parameter	Description	Value	Targeted Moment
σ_s	Tfp innovation dispersion	0.29	Labor Dynamism
ϕ_L	Adjustment cost	0.0032	Labor adjustment as fraction of revenue
ϵ/σ	Super-elasticity	0.6	Labor–sales regression
μ_F	Log fixed cost mean	-3.15	Entry rate
σ_F	Log fixed cost dispersion	1.65	Average size exiting 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 30: The response of the baseline quantitative model to an MIT shock



D Solution method

D.1 Simple model

Observe that in the static version of the model (without any labor adjustment costs) the firm's problem is

$$\pi(z) = \max_{y \geqslant 0} \left[\Upsilon'(y/Y)Dy - W\frac{y}{z} \right] \tag{D.1}$$

$$= \max_{y \geqslant 0} \left[\Upsilon'(y/Y)Dy/YY - W\frac{y/YY}{z} \right] \tag{D.2}$$

$$= \max_{q \geqslant 0} \left[\Upsilon'(q)DqY - W\frac{qY}{z} \right] \tag{D.3}$$

$$= DY \max_{q \geqslant 0} \left[\Upsilon'(q)q - \frac{W}{D}\frac{q}{z} \right] \tag{D.4}$$

$$\tag{D.5}$$

The optimal choice of q is then the same as the optimal solution to a modified problem

$$\tilde{\pi}(z) = \max_{q \geqslant 0} \left[\Upsilon'(q)q - Aq \right] \tag{D.6}$$

$$\tag{D.7}$$

where $A = \frac{W}{D}z$. Solving for the equilibrium in this model proceeds in two steps:

1. Solve for $q(A)$ on a grid of values of A.
2. Using this policy function, find the value of $\Omega \equiv W/D$ such that

$$\int \Upsilon(q(\Omega z))dH(z) = 1$$

Given the value of Ω and policy q, we can find D, aggregate markups, and productivity. The household preferences then pin down output, labor supply, and the wage.

D.2 Quantitative model

The quantitative model is somewhat more complicated, as we cannot solve an equivalent problem that depends on only one aggregate variable. 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:

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

A change in A thus only affects the location of the Pareto distribution (up to rescaling). I show an example of this kind of shift in Figure 31

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?

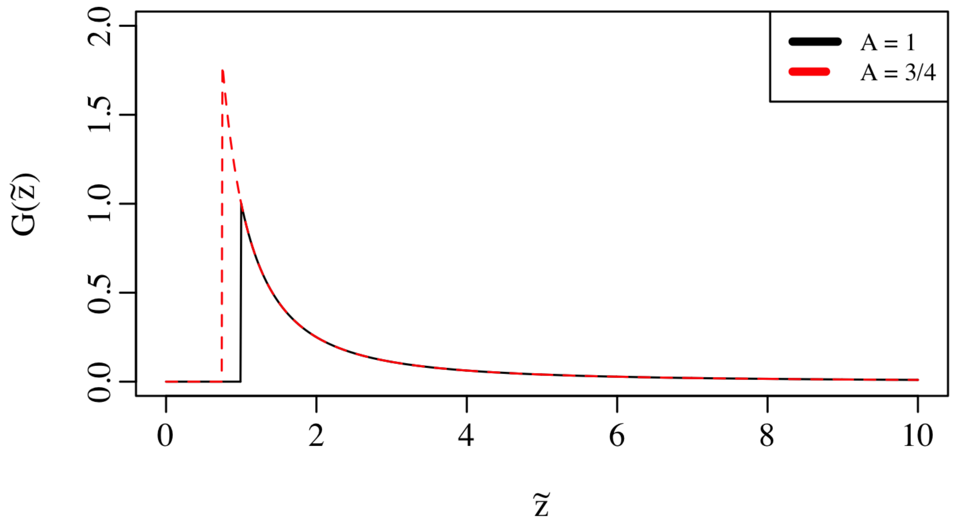
$$\mathcal{M} = \frac{\int_0^\infty \mu(q(z))\frac{q(z)}{z} dG(z)}{\int_0^\infty \frac{q(z)}{z} dG(z)}$$

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:

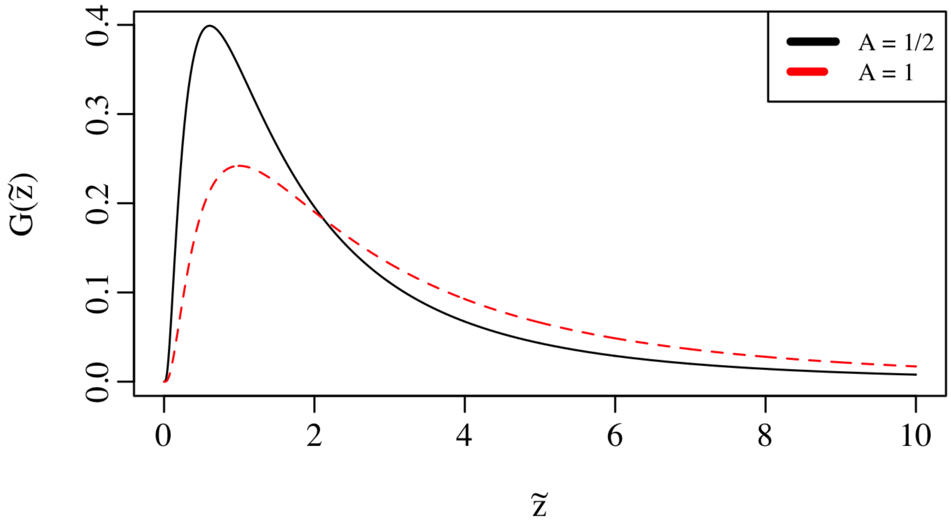
$$\mathbb{E}[(\tilde{z} - \mathbb{E}(\tilde{z}))^2] = \exp(\sigma^2) - 1) \exp(2(\log A + \mu) + \sigma^2)$$

Figure 31: A change of variables under the Pareto assumption



An increase in $\log A$ then increases both the mean and variance of $\tilde{z}$. Figure 32 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 32: A change of variables under the log-normal assumption



F Entry cost shock

A natural shock to study in this environment is a shock to the cost of entry. This shock has qualitatively similar effects to the shock to the mass of potential entrants that I discuss in the paper. The key difference is that a shock to the cost of entry, through a selection effect, increases the average productivity of entrants. This implies that the employment share among entering and young firms does not fall by as much as it does in the entry mass shock.

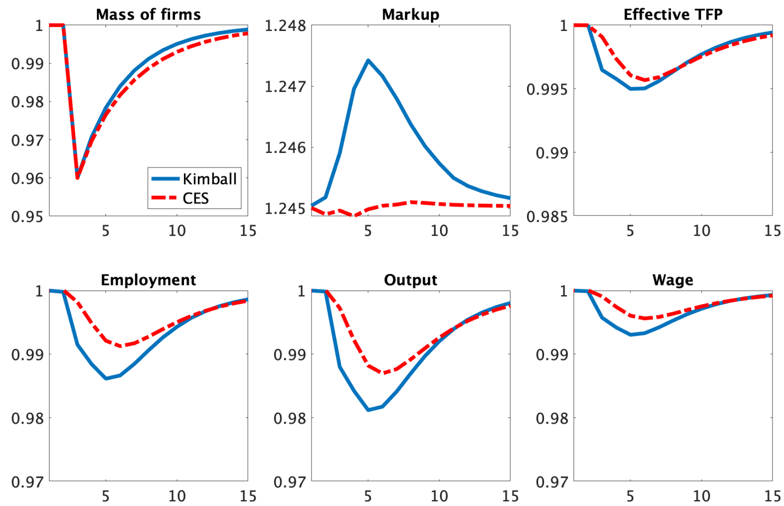
Figure 33 summarizes the effects of a shock to the cost of entry in the Kimball and CES models. The figure looks qualitatively similar to the behavior of aggregates following a shock to the mass of potential entrants.

The main difference between the shock to the mass of potential entrants and the shock to the cost of entry turns out to be the selection effect. Recall that the optimal policy of potential entrants is to enter if and only if

$$V_E(\phi) \geqslant c_E$$

V_E is increasing in the signal ϕ, because ϕ is positively correlated with future productivity. This implies that there is a cutoff rule: firms enter if and only if $\phi \geqslant \tilde{\phi}$. A rise in the cost of entry c_E implies that only firms with higher values of the signal ϕ enter. A fall in the mass of entrants has the opposite implication: it increases V_E for every value of ϕ and so actually leads $\tilde{\phi}$ to fall.

Figure 33: Shock to the cost of entry



The rise in the average productivity of entrants is evident in the path of the share of entrants in aggregate employment following this shock. In spite of the fact that the entry rate falls from 11.5% to under 7%, the share of employment at entering firms only falls from 5.5% to 5.2%. This paltry drop is due to the large increase in the average signal of entrants, whose value rises by 20%. Figure 34 depicts the paths of these variables.

The entry mass shock, on the other hand, reduces the average size of entrants as well. Figure 35 depicts the path of these variables following the entry mass shock. As it shows, the entry mass shock reduces the average productivity of the entrant firms.

Figure 34: Entrants following the entry cost shock

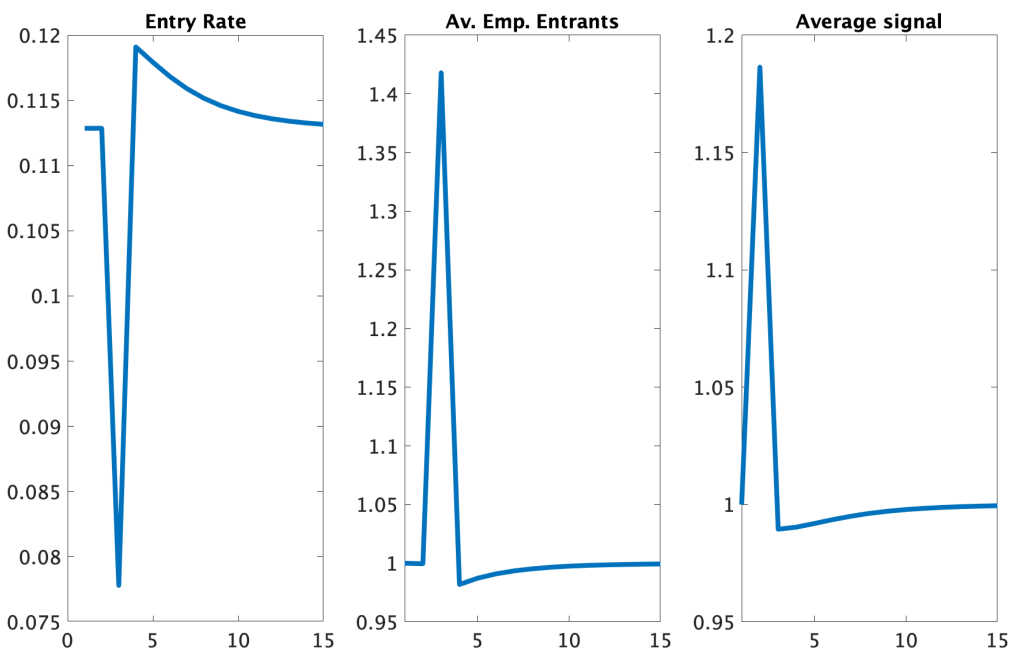
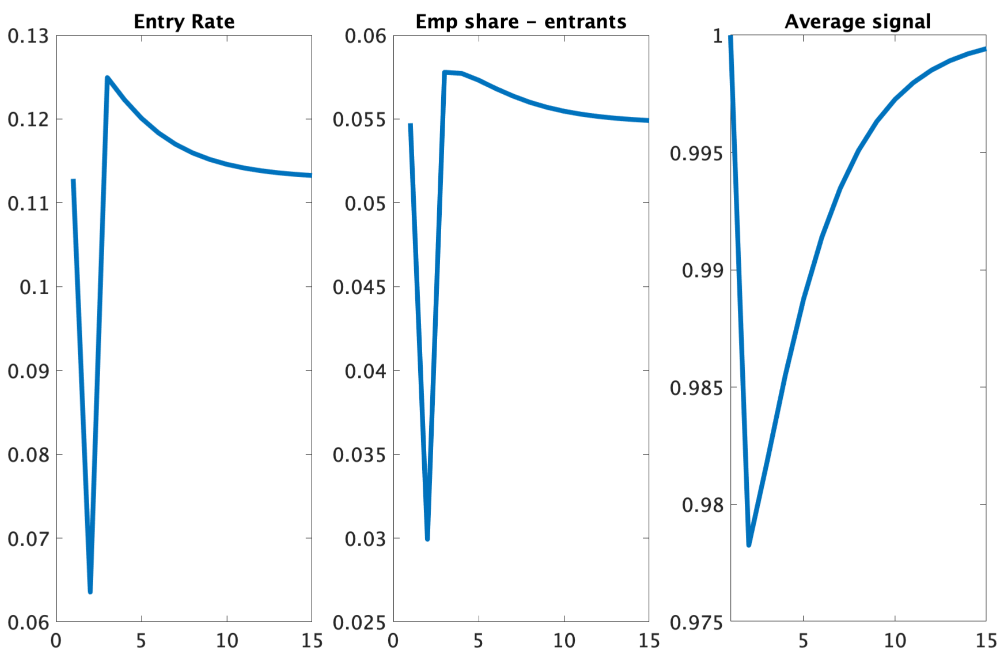


Figure 35: Entrants following the entry mass shock



G Financial shock

One cause of a fall in entry could be a rise in the cost of borrowing. To study this more formally, I introduce financial frictions and study a shock to the cost of obtaining financing. Following Gilchrist et al. (2017), I assume that firms face a cost of issuing equity.

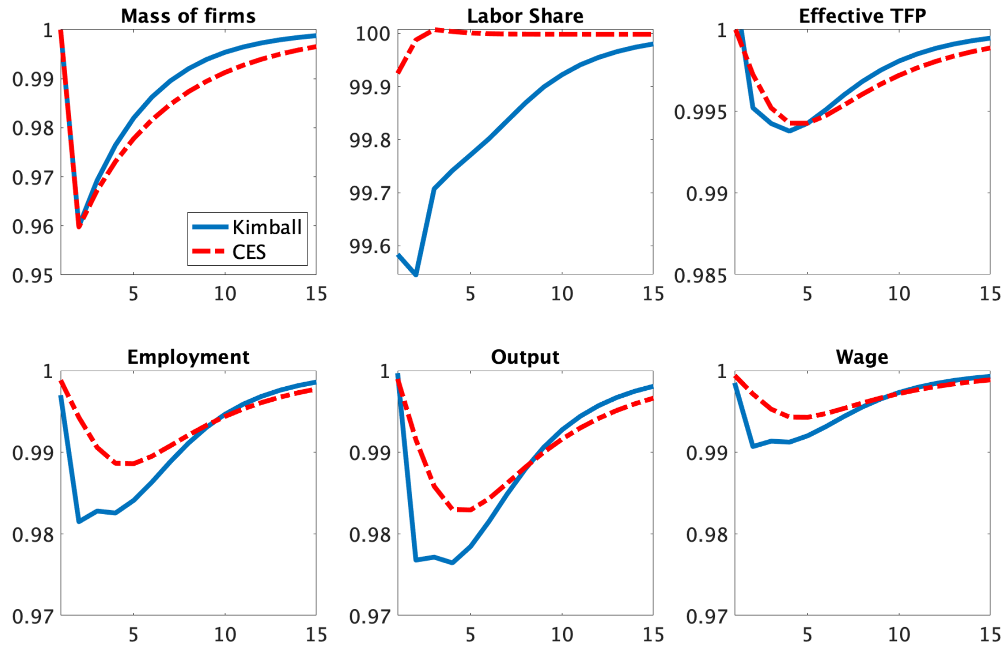
Let φ denote the cost of issuing equity (i.e., of having negative profits). Moreover, suppose that there is a deterministic component to the fixed cost $\bar{c}_F$[15]. The firms' flow value is:

$$F(z, L, L') = \begin{cases} py - WL' - c_F - c(L, L') & \text{if } F > 0 \\ \frac{1}{1-\varphi}(py - WL' - c_F - c(L, L')) & \text{if } F < 0 \end{cases}$$

I also assume that firms must borrow to cover their entry costs as well, and so the entry cost becomes $\frac{1}{1-\varphi}c_E$. A shock to ϕ then acts to both (1) reduce the size of small firms and (2) reduce entry. Figure G depicts the results of this experiment. I scale the shock so that the initial path of the mass of firms is the same (falls by 4%) in both models.

[15]There is still some component of the fixed cost that is random. This is useful for targeting the average size of exiting firms.

Figure 36: Financial shock



H Stochastic Discount Factor

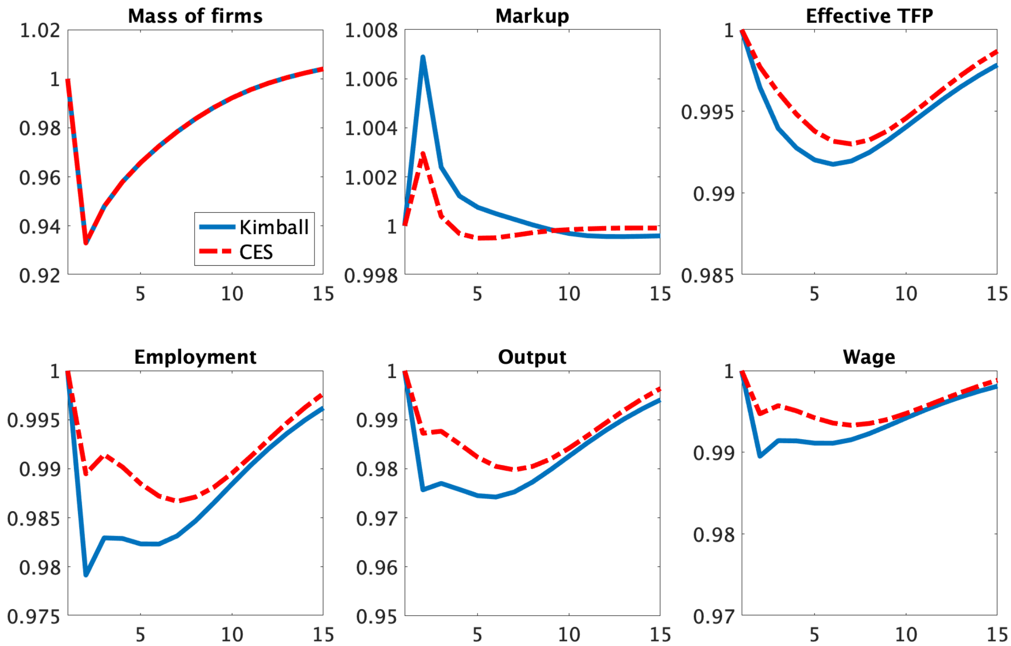
H.1 Shock to entry

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

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

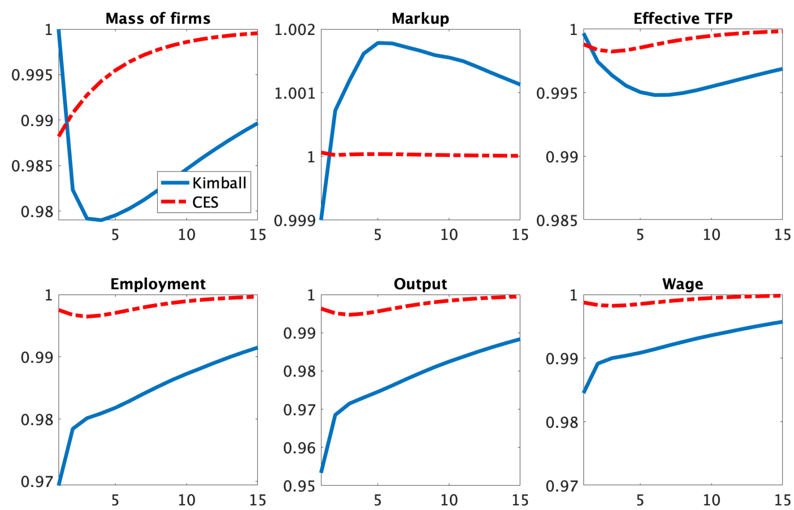
I set $\gamma = 1$. The impulse response functions for the Kimball and CES economies to this shock are depicted in Figure 37. As they show, the variable SDF increases the persistence of the effects of the shock and the significance of the variable markups channel. The fall in the stochastic discount 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 37: Impulse response to an entry shock; variable stochastic discount factor



H.2 TFP shock

The mass of entrants fluctuates strongly with aggregate TFP shocks in a model with a pro–cyclical stochastic discount factor.


Mass of firms
Markup
Effective TFP
Employment
Output
Wage
Kimball
CES
1
0.995
0.99
0.985
0.98
1.002
1.001
0.999
0.97
0.96
0.95
5
10
15