

Entry, Variable Markups, and Business Cycles

[Click here for most recent version](#)

William Gamber*

October 9, 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 model. The key mechanism in the model is that declining entry leads the market shares of large incumbent businesses to rise, and in response they increase their markups and reduce their employment. I present evidence for this relationship, showing in a panel dataset of large firms that the within-firm elasticity of the markup to relative sales is 35%. I then study shocks to entry in a model that is consistent with this elasticity, finding that a fall in entry can lead to a significant contraction in employment. Much of this effect is due to entrants’ impact on incumbent firms. A shock to entry that replicates the decline in the number of businesses during the Great Recession generates a prolonged 5 percent fall in employment in the model. Finally, I show that the increasing correlation between market shares and markups over the last 30 years implies that the effect of entry on the business cycle is stronger today than it was in 1990.

*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, Seb Graves, James Graham, Venky Venkateswaran, and Josh 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 dramatically, falling by more than 35% relative to its peak in the mid 2000s and remaining depressed through 2014, 5 years after the end of the recession.¹ In this paper, I study how fluctuations in the creation of new businesses amplify recessionary contractions in employment and output. I show that a key part of the propagation of the fall in business formation to the economy at large is the response of incumbent firms to the reduction in competition that they face from entrants.

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 a new 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 the fall in entry during the Great Recession led markups among large incumbent businesses to increase and significantly contributed to the contraction in aggregate employment.

There is a recent literature studying the effects of the decline in entry during the Great Recession. [Siemer \(2014\)](#), [Moreira \(2017\)](#), and [Clementi and Palazzo \(2016\)](#) all find that because entrants are small relative to incumbents, their absence does not have large immediate effects on aggregate employment. In contrast to these papers, I find that a contraction in entry has large and immediate effects on aggregate employment and output. The key difference between my work and theirs is that in my model, large firms increase their markups as their market shares rise. This means that, even though entrants are small, a shock to entry causes a large and persistent contraction in aggregate employment because in response to the decline in competition coming from entrants, large incumbent firms increase their markups and reduce their employment. This mechanism doubles the effects of a shock to entry on aggregate employment relative to models in which markups do not vary with market share.

I begin by providing empirical evidence that large firms increase their markups as their market shares rise. I show in a panel dataset of large firms that, controlling for firm heterogeneity, the typical firm increases its markup significantly as its market share grows. The positive relationship between market share and markups means that firms become less employment-intensive as they grow relative to their competitors, and so a shock that increases market concentration reduces employment.

I then study business cycles in a general equilibrium firm dynamics model that is consistent with this finding. In the model, firms face a demand curve with an elasticity of demand that declines with relative size, which implies that their markups rise as

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

they grow. The model features a firm life cycle and labor adjustment costs, and it is consistent with the joint distribution of firm age and size, the distribution of hiring decisions, and the within-firm relationship between markups and market share.

I show in the model that a temporary decline in entry has large and persistent effects on aggregate employment. A fall in entry increases the market shares of incumbent businesses and leads them to increase their markups, produce less, and reduce employment. These effects are most pronounced for the most productive firms, leading aggregate productivity to fall as well. Because of these changes, aggregate output and employment decline persistently.

Much of the contraction in output and employment is due to the variable markups of large firms. 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 cannot rationalize the within-firm relationship between markups and market share that I document in data – it 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.

I then turn to two applications of this model. First, I study the persistent decline in the number of establishments during the Great Recession. A shock that replicates the decline in the number of establishments relative to trend over the period from 2007-2014 leads employment to decline by 5 percent, recovering to trend only in 2020. In the model, half of that contraction is due to variable markups. This exercise suggests that policies to extend credit to potential new businesses or to help cover the fixed costs of small incumbents 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 markups and market share has grown significantly since 1985; my estimates imply that the elasticity of the markup to revenue has more than doubled over the past 30 years. I account for this increase in the model with an increase in the rate at which the elasticity of demand changes with relative size. I show that this increase implies that entry fluctuations have larger effects on aggregate employment today than they used to. 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

Entry and business cycles

There is a long literature studying the role of entry in business cycle models. My approach is unique in that it incorporates both variable markups and firm heterogeneity into a general equilibrium business cycle framework.

An older literature studies the effects of pro-cyclical entry on markups in homogeneous firm models. [Jaimovich and Floetotto \(2008\)](#) and [Bilbiie, Gironi and Melitz \(2012\)](#) both find significant effects of entry on aggregate quantities through its effects on markups and productivity. My paper introduces heterogeneity in firm size to these models. This is important for two reasons: First, entering firms are significantly smaller on average than incumbent firms, which limits the effects of entry on the market shares of incumbents. Second, as [Arkolakis et al. \(2019\)](#) document, even when entrants are the same size as incumbents, introducing heterogeneity into variable markups models tends to reduce the effects of entry on aggregates.

Another literature studies the effects of entry on output, taking into account that entrant firms are smaller than average than incumbents. [Siemer \(2014\)](#) and [Moreira \(2017\)](#) both document that young firms start small and grow slowly. 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 and grow slowly because of persistence in productivity and hiring costs. 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.

Heterogeneous firms, markups, and entry

There is a literature studying the effects of entry on markups. [Edmond, Midrigan and Xu \(2018\)](#) and [Arkolakis et al. \(2019\)](#) find that entry has little effect on the cost-weighted markup. In their model, small firms are most exposed to competition, and so while a fall in entry increases the markups of all firms, it also reallocates output away from high markup to low markup firms. There are two key differences between my model and theirs: first, we use different productivity distributions, and second, my paper includes adjustment costs. These two changes prevent the reallocation of output to low markup firms from undoing the firm-level increase in markups.

Causal evidence for entry’s effects on incumbents

This paper also presents a theory that rationalizes causal empirical 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 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.

Secular trends in markups and firm dynamism

A fact that I document in the paper is 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 [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. While this approach prevents me from measuring the level of markups, I find interesting results about their relationship

to firm size. 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.

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 \(2018\)](#), [Baqae 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.

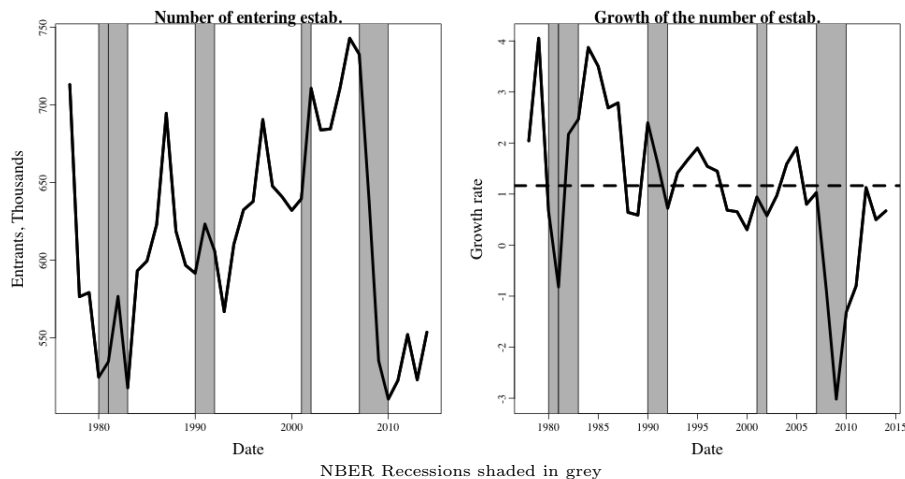
2 Background: entry over the business cycle

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

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



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.² 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.

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

²Note 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 2: Entry and exit of establishments in BDS

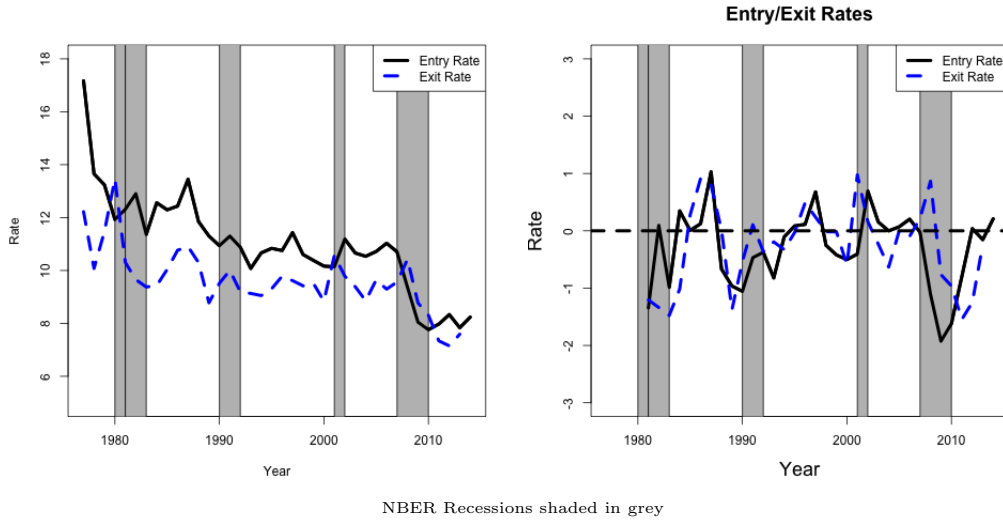


Table 1: Entrants relative to the whole economy, 1985–2014

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%

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 .³ 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:

³It 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.

$$Y = Q(A; K, L) \quad (3.1)$$

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 functions⁴.

$$Y = AK^\theta L^\alpha \quad (3.2)$$

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:

$$\min_{V, K} WL + rK + F \quad (3.3)$$

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

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

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

Multiplying both sides by L gives

$$WL = \alpha\lambda Y \quad (3.6)$$

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.

$$WL = \alpha \frac{PY}{\mu} \quad (3.7)$$

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:

$$\log WL = \log \alpha + \log PY - \log \mu \quad (3.8)$$

Consider the following regression:

⁴See [De Loecker and Eeckhout \(2017\)](#) for a more complete discussion of this approach.

$$\log WL = \tilde{\alpha} + \beta \log PY + \epsilon \quad (3.9)$$

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

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

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.⁵

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.

⁵This 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:

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

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 (0.0009***)	0.6275 (0.0016***)	0.356 (0.0137***)
$\log XLR$	0.8983 (0.003***)	0.6716 (0.007***)	0.4266 (0.007***)
$\log COGS$	0.9263 (0.0007***)	0.783 (0.002***)	0.654 (0.002***)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry \times Year	Firm + Industry \times Year	Industry \times 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.⁶ 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

⁶The 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 (0.0009***)	0.3735 (0.0016***)	0.644 (0.0137***)
$\log XLR$	0.1017 (0.003***)	0.3284 (0.007***)	0.5737 (0.007***)
$\log COGS$	0.0737 (0.0007***)	0.217 (0.002***)	0.346 (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.

Relaxing assumptions

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.

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.

Table 5: Variable input use and relative size over time

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

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 estimates⁷ 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 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

⁷This is not surprising given the sparsity of data available for that measure.

Figure 4: Variable Input–Revenue Relationship, Rolling Windows

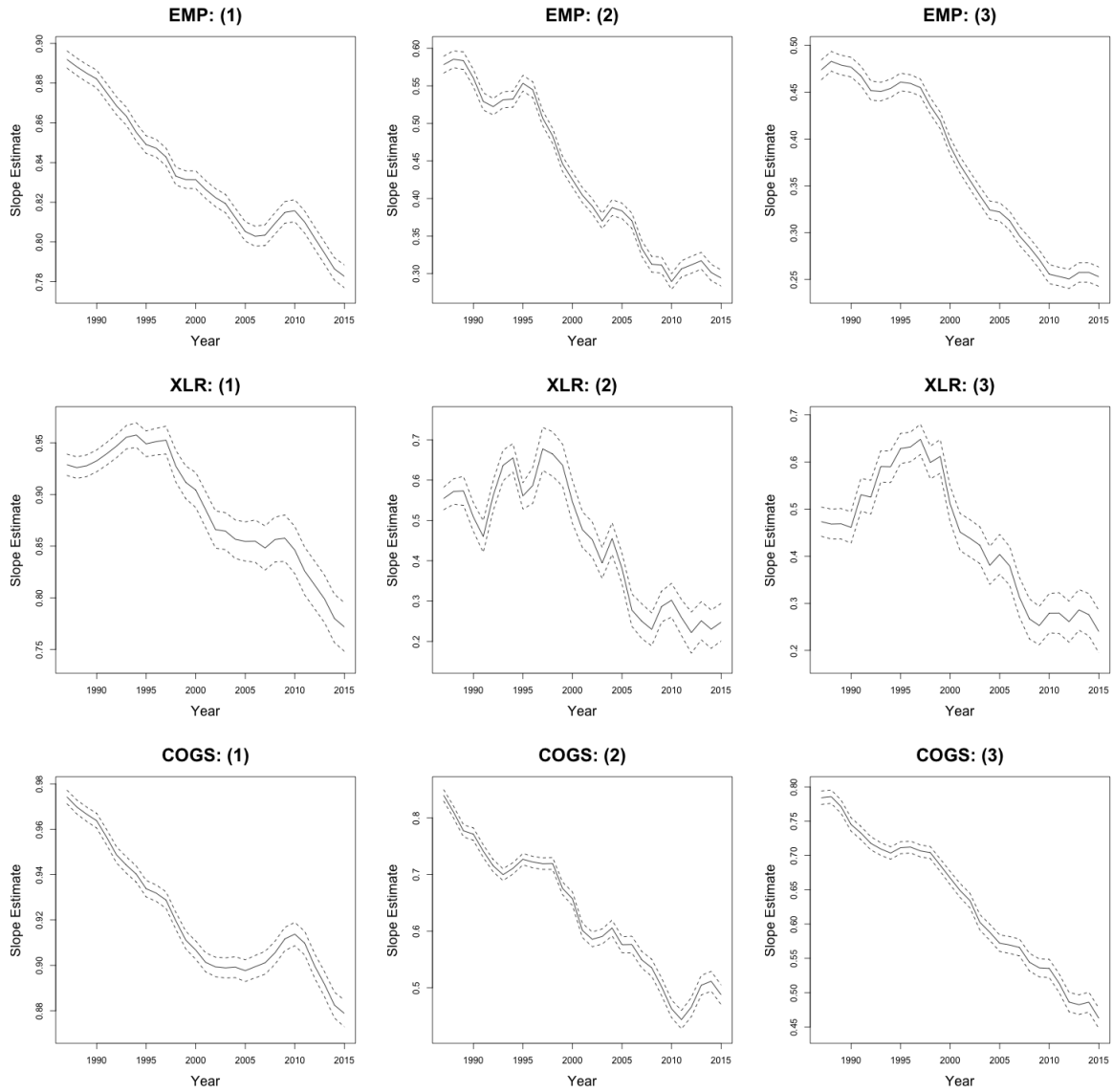


Table 6: Dynamism in Compustat, 2010

Measure	Reallocation
<i>EMP</i>	6.17 %
<i>XLR</i>	7.24 %
<i>SALE</i>	14.15 %

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:

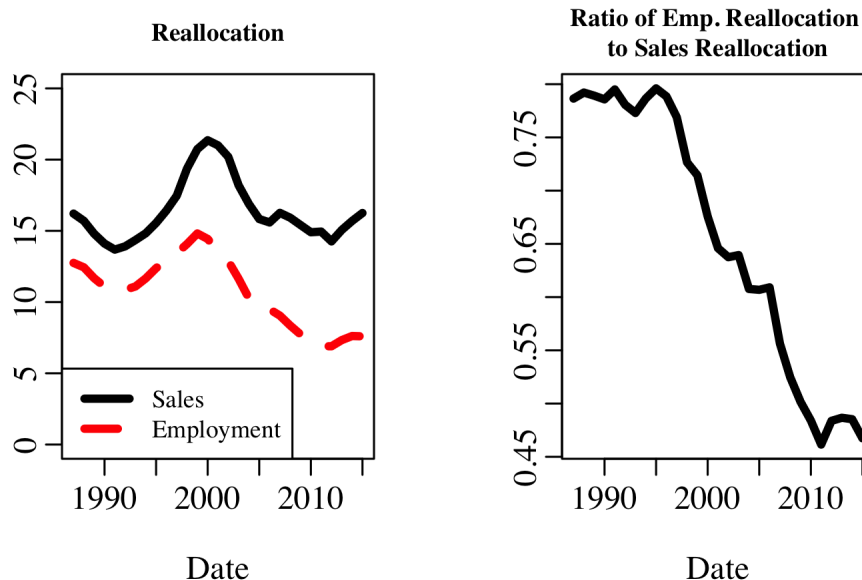
$$\underbrace{\text{Var}(\Delta \log PY)}_{\text{Sales reallocation}} = \underbrace{\text{Var}(\Delta \log WL)}_{\text{Employment reallocation}} + \underbrace{\text{Var}(\Delta \log \mu) + 2\text{Cov}(\Delta \log \mu, \Delta \log L)}_{\text{Markup variation}} \quad (3.12)$$

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 and its covariance with employment growth.

These measures have not been stable over time. As emphasized in [Decker et al.](#)

Figure 5: Employment and Sales Dynamism



(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. Previous literature on firm entry over the business cycle has emphasized adjustment frictions, and so the firms in this model face labor adjustment costs. 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:

$$C_t \leq W_t L_t + \Pi_t \quad (4.2)$$

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

$$W_t = -\frac{u_{L,t}}{u_{C,t}} \quad (4.3)$$

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:

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

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 $\Upsilon(q)$:

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

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

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

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:

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

where D_t is termed the “demand index” and is defined as

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

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

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

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 :

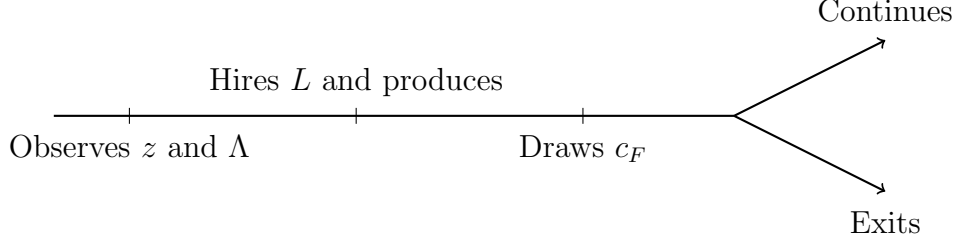
$$F(L; z) = zL^\eta \quad (4.10)$$

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. If it chooses to continue producing in the next period, it faces a probability γ of being exogenously forced to exit and then continues to the next period.

Let Λ summarize aggregate states that are relevant to each establishment. The recursive problem of an incumbent establishment who employed L employees last period,

Figure 6: Timing for incumbent establishments



has productivity z and has paid fixed cost ϕ_F is:

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

$$\tilde{V}(L, z, c_F; \Lambda) = -c_F + \beta(1 - \gamma) \mathbb{E} \left[V(L, z'; \Lambda) | z \right] \quad (4.12)$$

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

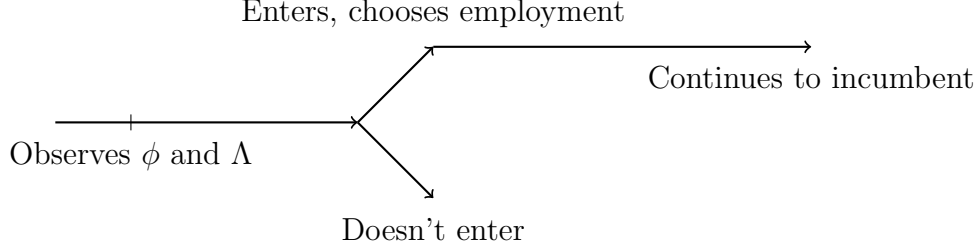
$$y \leq zL \quad (4.14)$$

Stochastic discount factor

In the deterministic steady state, the firm discounts future streams 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 I, 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

Figure 7: Timing for potential entrants



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.

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:

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

The optimal policy of the potential entrant is to enter if and only if $c_E \leq V_E(\phi)$. 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 and
4. consumption C and labor supply L satisfy the household first order condition
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 economy⁸. Consider the aggregate production function, where Z_t denotes *aggregate productivity*:

$$Y_t = Z_t L_t \quad (4.16)$$

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

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

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 firms restrict their output.

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

$$\mathcal{M}_t = \frac{Y_t}{W_t L_t} \quad (4.18)$$

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:

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

⁸Though, solving the model still requires approximating the value function of the firms. See Appendix E.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.

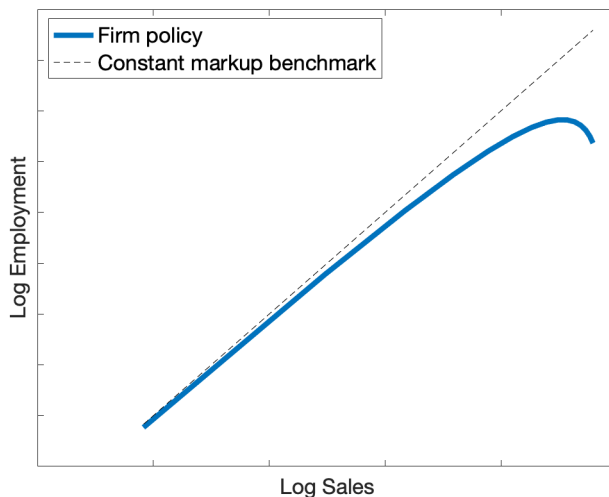
I calibrate the model to the behavior of establishments. Establishments are more likely to represent unique products and thus might be better thought of as the relevant unit of competition for this model. As Table 1 shows, entering establishments are larger relative to incumbent establishments than entering 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 D , 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 external data on adjustment costs as a fraction of revenues, 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.

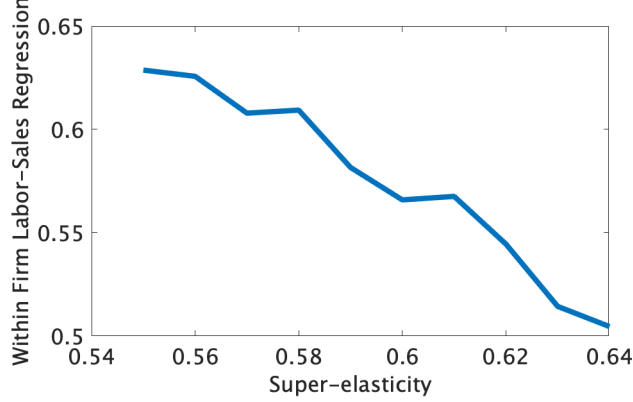
The sample I use in Compustat covers about 1% of firms and 30% of U.S. non–farm payroll. In my simulated method of moments estimation procedure, I simulate a sample of firms in the model and then estimate the regression on a subsample of the top 1% of firms by sales in the model economy. This procedure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



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

These imply a labor supply curve:

$$\psi L_t^\nu = W_t \quad (5.2)$$

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:

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

Parameterization

I exogenously set some parameters and then jointly choose seven of them to target important moments. The model period is one year, so I set $\beta = 0.96$ and $\rho_z = 0.53$. These parameter choices are summarized in Table 7. I then estimate the remaining parameters using simulated method of moments, jointly choosing productivity innovation dispersion σ_z , the adjustment cost parameter ϕ_L , the demand parameters σ and ϵ , the average fixed cost μ_F , fixed cost dispersion, σ_F , and the Pareto parameter for the distribution of entrant signals ξ . To simplify the calibration procedure, I set the sunk cost of entry equal to the expected value of the distribution of fixed costs of production.

Figure 10: The distribution of the signal and productivity

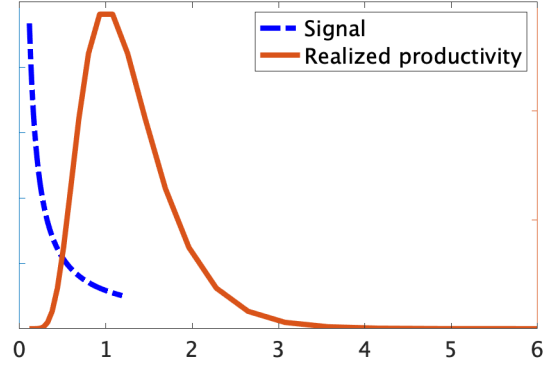


Table 7: Pre-set parameters

Parameter	Description	Value	Source/Target
β	Discount factor	0.96	Annual model
σ_z	Idiosyncratic tfp innovation variance	0.53	Cooper and Haltiwanger (2002)
σ	Kimball demand elasticity	10	
γ	Exogenous exit rate	1.5%	
M	Mass of entrants	1	Normalization
ν	Inverse Frisch Elasticity	0.5	CP
δ	Job separation rate	0.19	

DLEU: [De Loecker and Eeckhout \(2017\)](#), CP: [Clementi and Palazzo \(2016\)](#)

Table 8: 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

While each of these parameters affects several moments in the model, they each intuitively correspond to one particular moment. Productivity innovation dispersion affects the cross-sectional variance of firm-level log employment growth, which I estimate to be 7.5% in Compustat. The adjustment cost affects the average size of the total adjustment cost as a fraction of revenues, which [Bloom \(2009\)](#) estimates to be 2.1%. The super-elasticity directly 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. The average fixed cost affects the exit rate and thus the steady state entry rate. Fixed cost dispersion affects the average size of exiting firms; if it is high, the average exiting firm will look like the average firm in the economy, otherwise, there is a high degree of selection on exit and exiting firms will on average be small. Finally, the Pareto parameter for the entrant signal affects the relative size of entering firms. Table 8 summarizes the parameter choices as well as their identifying moments in the model and in the data.

The model performs well along a number of targeted and untargeted moments. Figure 9 summarizes the model’s fit. As in the data, the model generates a wedge between labor and sales dynamism. The wedge between these two numbers is 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.

Table 9: Calibration Targets & Model Fit

Moment	Target	Source	Model moment
Labor dynamism	7.5%	Compustat	4.97%
Adjustment cost size	2.1 %	Bloom (2009)	1.81%
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%
Share of employment at young firms	30%	BDS	37.03%
Sales dynamism	15%	Compustat	14.21%

DLEU: De Loecker et al (2019), CP: Clementi and Palazzo (2016)

Untargeted moments below line

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 17, estimates of the superelasticity using microdata tend to be below 1. My estimates are closest to [Amiti, Itskhoki and Konings \(2019\)](#), [Berger and Vavra \(2019\)](#), and [Gopinath, Itskhoki and Rigobon \(2010\)](#), 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. 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 0.05, close to the regression in compustat 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

Table 10: Selected parameterizations of [Klenow and Willis \(2016\)](#) demand

Paper	σ	ϵ	ϵ/σ
This paper	9	5.67	0.63
Edmond, Midrigan and Xu (2018)	10.18	1.4252	0.14
Amiti, Itskhoki and Konings (2019)	5	1.6	0.26
Berger and Vavra (2019)	5	2.35	0.47
Gopinath, Itskhoki and Rigobon (2010)	5	3	.6
Lindé and Trabandt (2019)			10
Smets and Wouters (2007)			12.55

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.

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

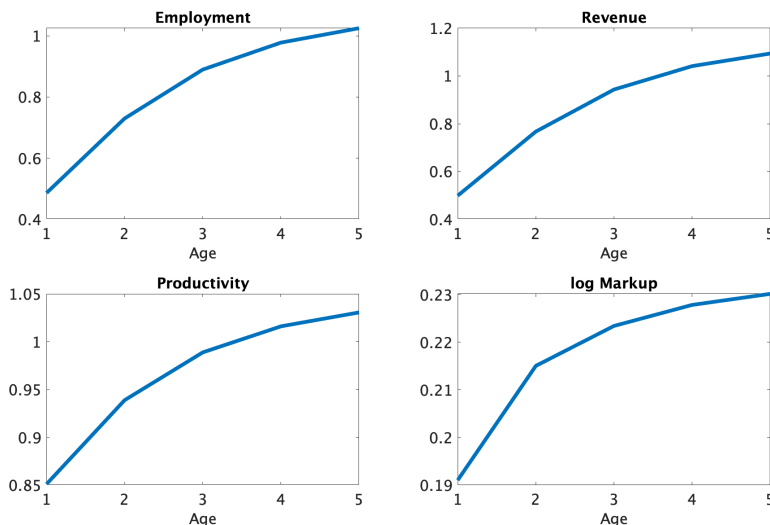
Firms in the model, as in the data, begin small and grow slowly. Figure 11 shows that employment and revenue at entering establishments are around 50% of the the average incumbent firm. They reach the size of the average firm by around age 5. The model achieves this in two ways: (1) the average productivity 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 4 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](#)

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



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.

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 relative 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.

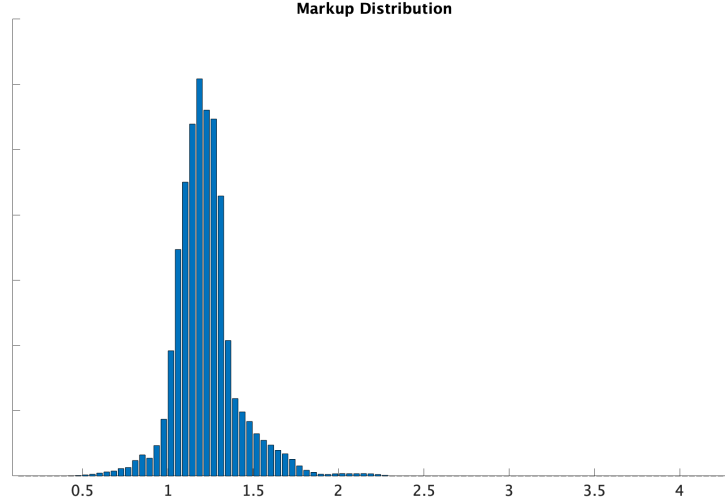


Figure 12: The cost-weighted distribution of markups

6 Business Cycles in the model

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 entrants⁹. 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 E.2.

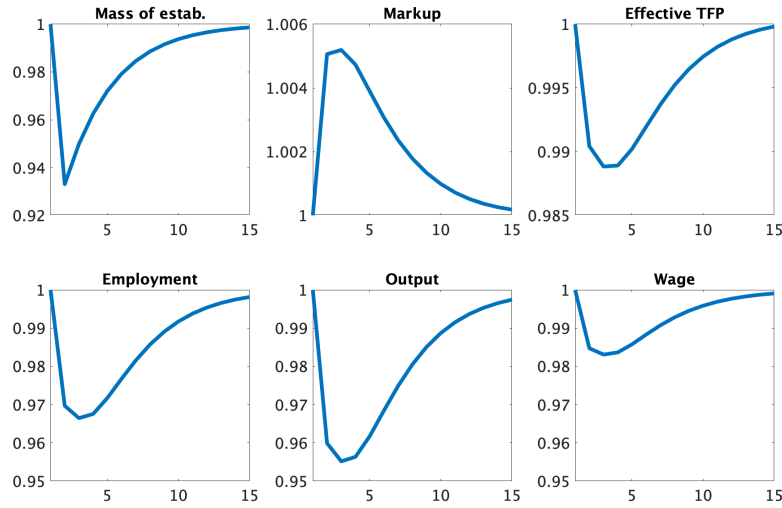
Entry shocks as financial shocks

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 G, I study a shock to the cost of entry. In Appendix H, I introduce a financial friction and study a shock to

⁹Why 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 G.

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



the cost of issuing equity, following [Gilchrist et al. \(2017\)](#).

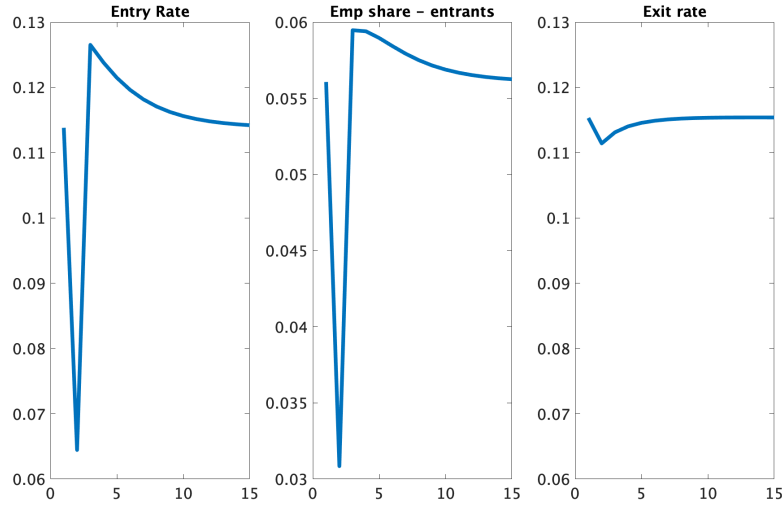
An entry shock

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 45 basis points. Since the labor share is the inverse of the average markup, it falls by 45 basis points. Effective TFP, the ratio of output to aggregate employment, falls endogenously by over 1 percent. Employment falls by 3 percent, and output falls by a bit more than 4 percent. The wage satisfies the household labor supply equation and falls by around 1.5 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

Figure 14: Entrants following the shock



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. This is exactly the “missing cohort effect”; since young firms slowly grow large, a missing generation of entrants has long term effects.

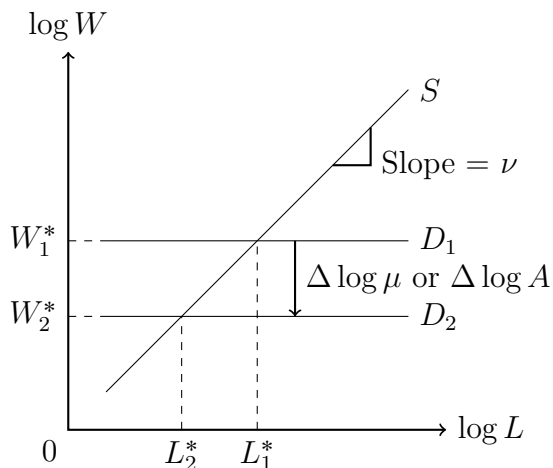
Markups and Productivity

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.

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



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:

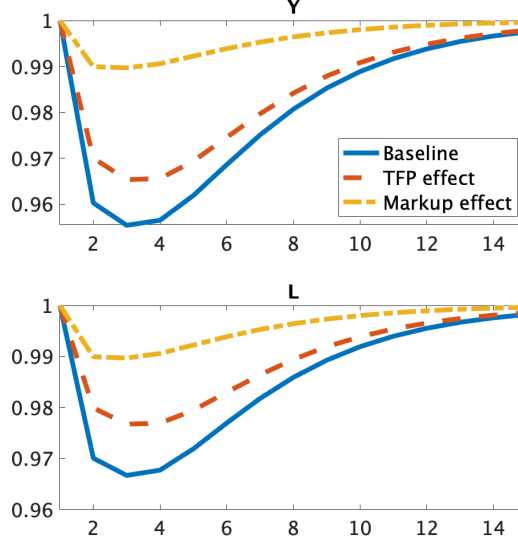
$$\log W_t = \log \psi + \nu \log L_t \quad (6.2)$$

$$\log W_t = \log A_t - \log \mu_t$$

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 45 basis points generates a 90 basis point 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, about 2/3 of the fall in employment is due to the fall in aggregate TFP and 1/3 is due to the rise in the cost–weighted markup.

Figure 16: Decomposition of entry shock



Aggregate TFP

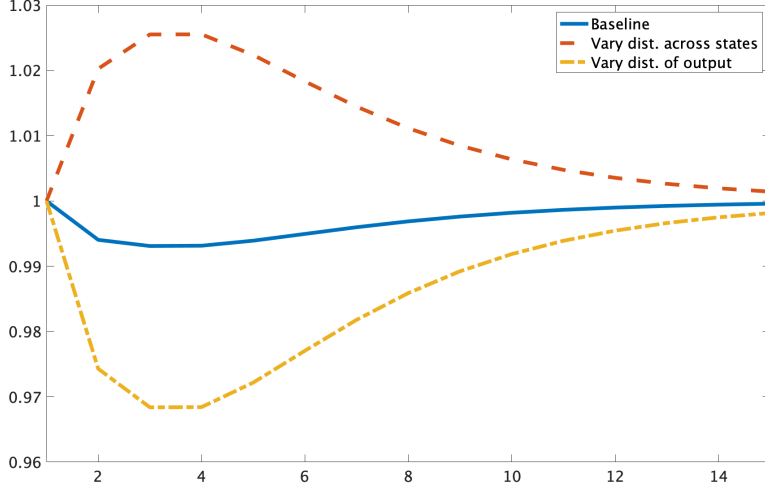
The decline in aggregate TFP accounts for a large portion of the fall in employment. To understand why aggregate TFP falls, I decompose its fluctuations into movements due to the change in the distribution of firms and those due to the change in the allocation of output across firms. Recall the definition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

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

In a purely 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.

Figure 17: A TFP decomposition



The cost weighted markup

The increase in the aggregate markup generates around one third of the contraction in employment. The relevant measure of the aggregate markup in this economy is the cost-weighted markup:

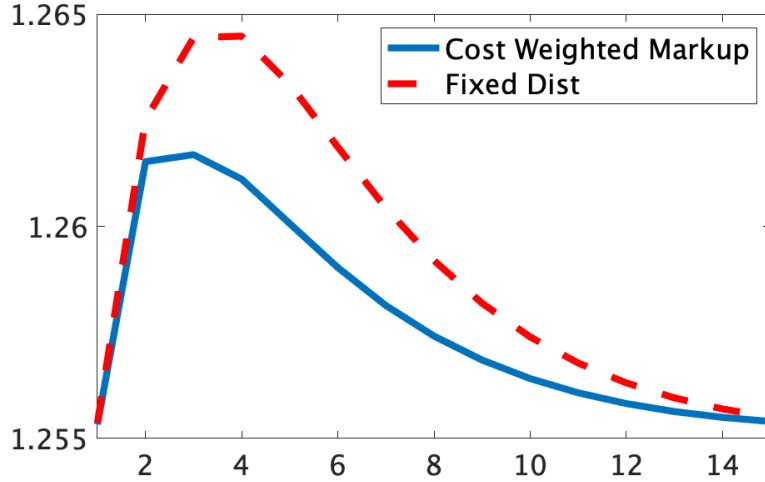
$$\mathcal{M}_t = \int \int \mu_t(z, L) \frac{\ell_t(z, L)}{L_t} d\Lambda_t(z, L) \quad (6.3)$$

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.

The reallocation effect is strong, undoing almost half of the increase in the cost-weighted 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. It is the cumulation of both of these two effects. 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

Figure 18: Decomposition of the response of markups



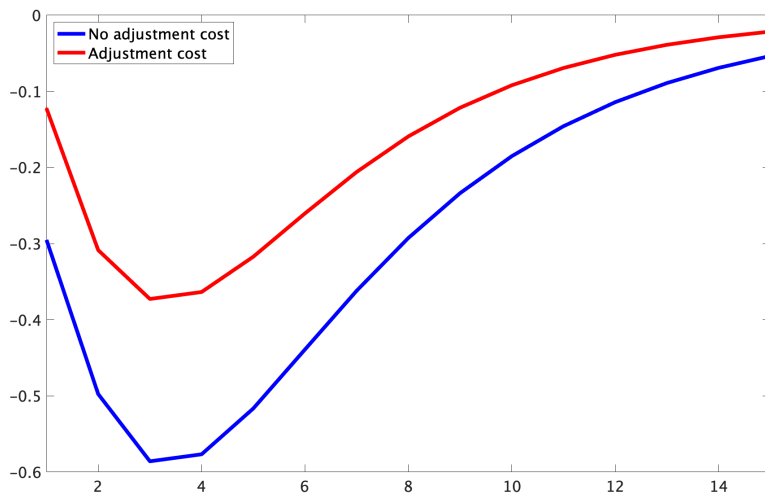
“missing deflation” during the last recession. Countercyclical markups lead inflation to rise. Any inflation measure that does not fully account for substitution effects would measure a rise in inflation that is larger than the cost-weighted markup increase.

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.

Figure 19: The role of the adjustment cost in reallocation



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 my model, there clearly are 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, undo some portion 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 that firms set prices statically (i.e., their production decisions today do not affect their future profits) 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 log-normal productivity. 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 F.

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

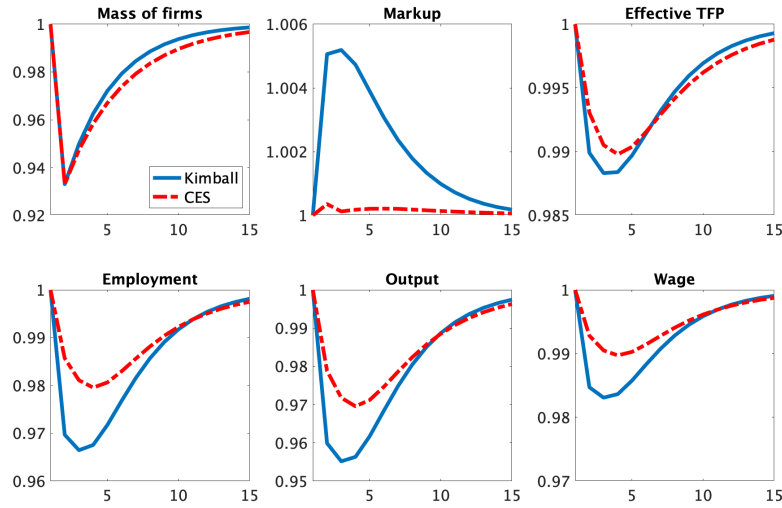
$$\Upsilon(q) = q^{\frac{\sigma-1}{\sigma}} \quad (6.4)$$

Without a labor adjustment cost, all firms would find it optimal to set markups equal to $\sigma/(\sigma - 1)$, so that markups no longer vary systematically with relative sales. The value of the elasticity of substitution I use in the CES calibration is $\sigma = 5.25$, which implies a frictionless markup of 1.235. The presence of adjustment costs raises the average markup in the economy to 1.25.

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. Employment and output in the CES model fall by only 61% and 67% as much as in the variable markups economy, respectively. Moreover, the decline in employment unfolds more quickly in the Kimball model, falling by more than twice as much than in the CES model in the first year after the shock.

The extra amplification in the Kimball model reflects both an immediate rise in the markup that is not present in the CES model and a larger fall in aggregate TFP in the Kimball model. In both models, the market shares of large incumbents rise following the decline in the entry rate. However, in the Kimball model, large incumbents raise their markups in response to the increase in their market shares. This leads them to reduce their employment, causing aggregate employment to fall. The reduction in the relative output of large firms in the Kimball model further reduces effective TFP relative to in the CES model.

Figure 20: Entry shock in the Kimball and CES models

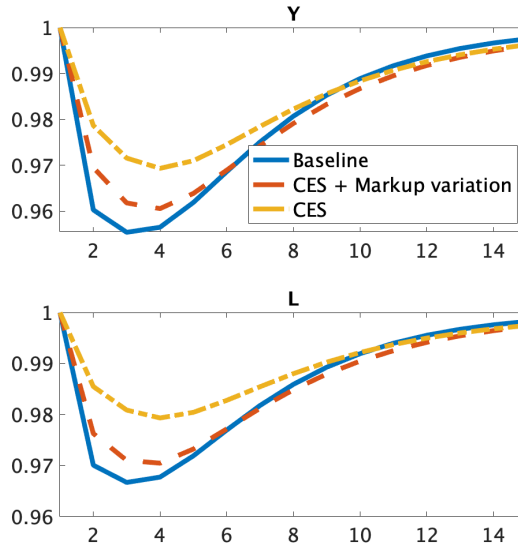


Decomposition of the role of variable markups

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. Thus, in spite of the fact that these two economies have the same markup and thus their love-of-variety effects ought to be similar, they still have significantly different paths for aggregate productivity.

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



7 Quantitative applications

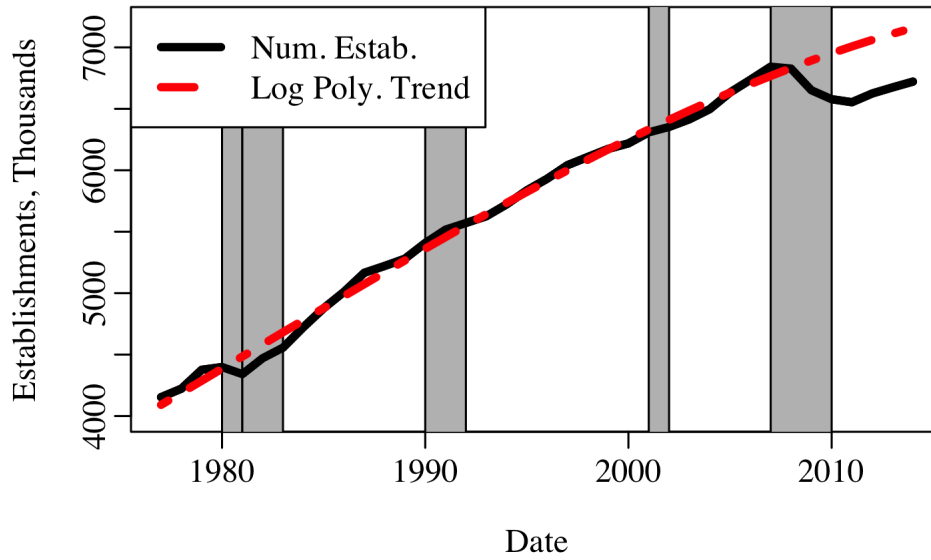
In this section, I study two applications of the model. In the first, I study the role of entry in the sharp contraction in employment during the Great Recession and its subsequent slow recovery. I show that an entry shock that reproduces the path of the mass of firms during the Great Recession leads 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.

Entrants are, on average, much smaller than incumbent firms, and so their share in total employment is lower than their entry and exit rates. 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.

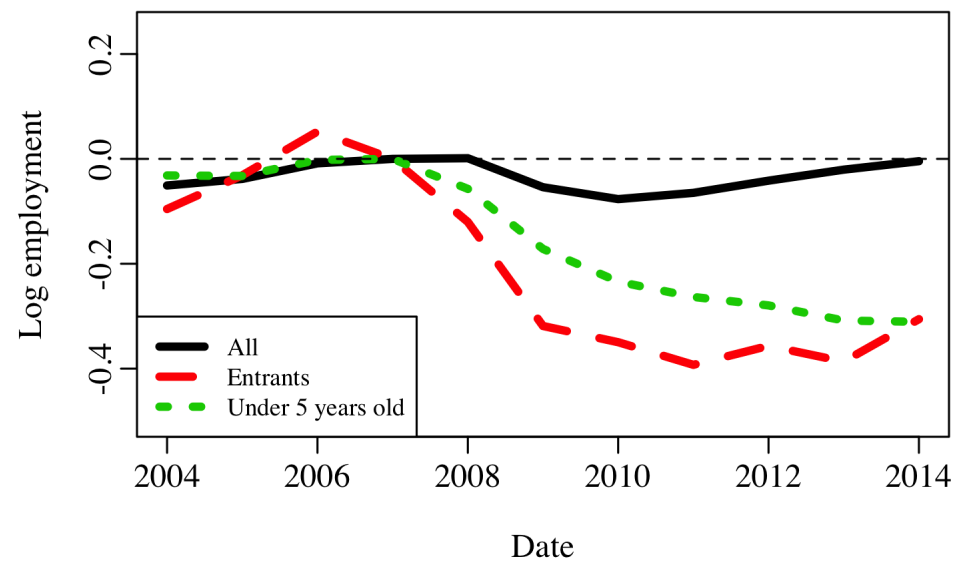
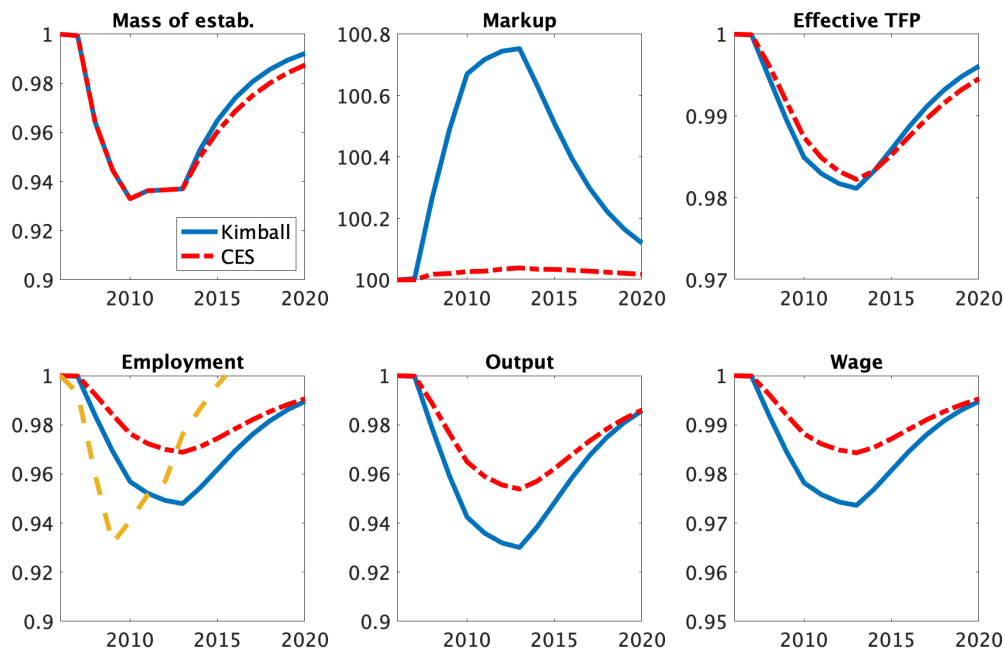


Figure 23: Age and employment among establishments

Figure 24: The Great Recession



The Great Recession in the model

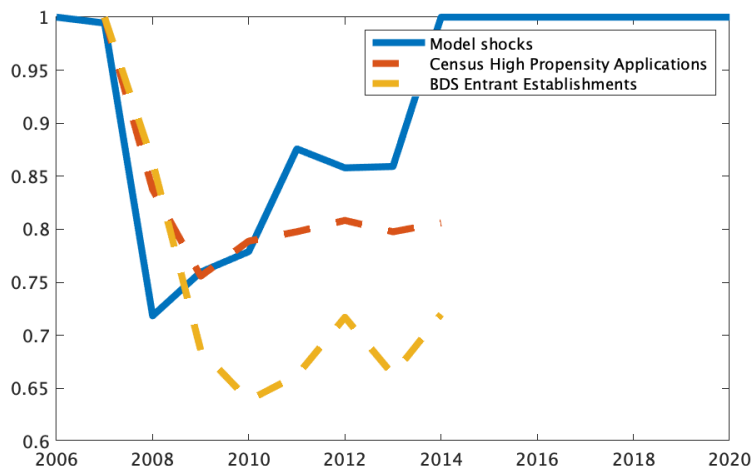
To understand the effects of the fall in entry during the Great Recession on markups, I feed in a sequence of shocks to the mass of potential entrants so that the path of the number of establishments in the model follows its path in data from 2007 to 2014. As before, I perform this experiment in both 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 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 classifies business applications as “high propensity” if they are likely to become businesses with a payroll¹⁰. As Figure 25 shows, the decline is slightly more than the fall in “high

¹⁰Includes applications (a) from a corporate entity, (b) that indicate they are hiring employees, pur-

Figure 25: The mass of potential entrants



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.

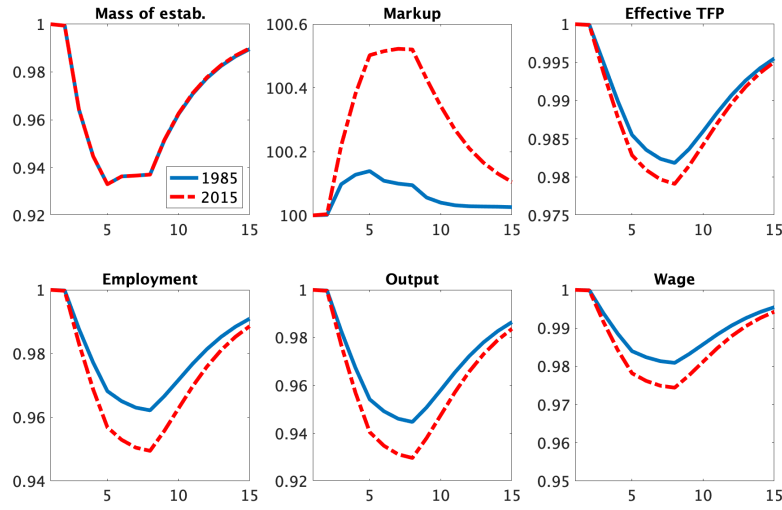
Table 11: Selected moments, 1985 vs 2015 calibration

Calibration	ϵ/σ	β_L	Labor rea./Sales rea.	Cost-weighted markup
1985	0.45	0.7663	60.0%	1.23
2015	0.75	0.463	24.3%	1.25

I choose the value of ϵ/σ to match the regression coefficient in 1985 of 0.786 and

chasing 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 26: Response to entry shock in 1985 and 2015



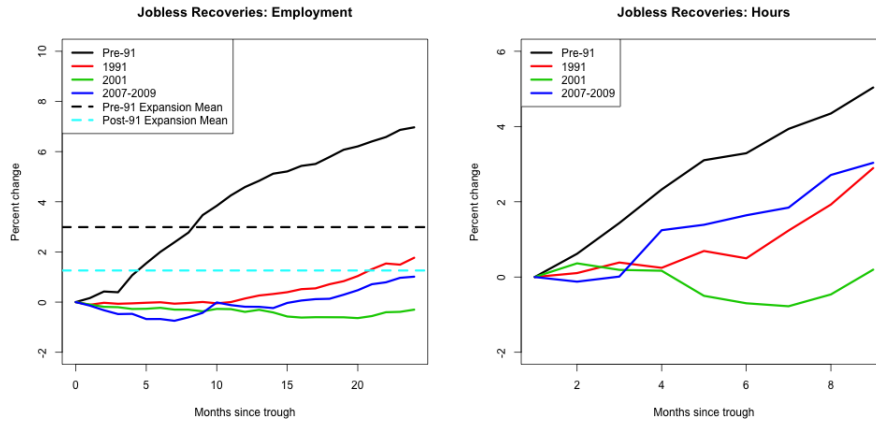
in 2015 of 0.486. As Table 11 shows, this generates a rise in the wedge between sales and labor dynamism, so that employment growth dispersion as a ratio of sales growth dispersion falls from 60% to 24.3%. This decline matches the decline of this ratio that I documented in Section 3. So, the higher covariance 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 can generate an increase in the cost-weighted markup from 1.23 to 1.25. 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 50 basis points and only gradually recovers in the 2015 calibration, but in the 1985 calibration, it rises by only 10 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 37% 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.

Figure 27: Jobless Recoveries



As it shows, employment recovers much more quickly following recessions before 1991 than after. Before 1991, employment recovered slightly more quickly than average following recessions, whereas after 1991, employment stagnated for almost a year before beginning to gradually recover. This timing coincides with the dramatic increase that I document in the markup–revenue relationship. During the Great Recession in particular, while other headwinds may have subsided, the anti-competitive effects of entry continued to buffer the employment recovery.

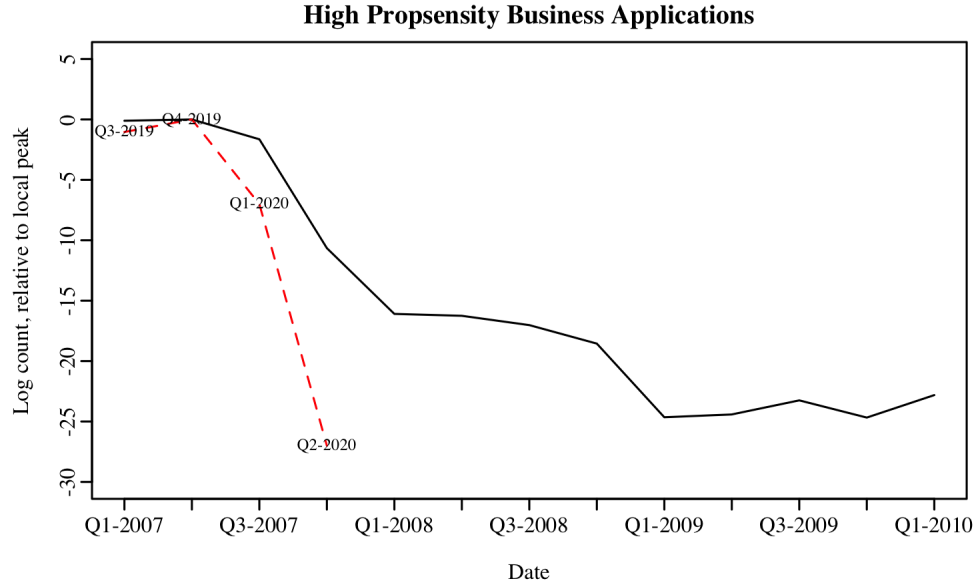
This exercise suggests that the rise in market power documented by [De Loecker and Eeckhout \(2017\)](#) and others might lead business cycles to become more volatile. As large firms’ markups become more responsive to their market shares, fluctuations in entry will increase the volatility of aggregate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

Firm creation during the 2020 Recession

As of the publication of this draft, there are no data on firm entry and exit for the first half of 2020. However, preliminary data suggest that declining business formation could contribute significantly to the large contraction in US employment.

Figure 28 plots the decline in “high-propensity business applications” during the Great Recession and in 2020. As it shows, high propensity business applications fell by 25 percent over the first 2 quarters of 2020. This steep fall is as large as the gradual decline of this measure over from the second quarter of 2007 through the first quarter of 2009. A sudden shock the size of the shock to entry during the Great Recession would lead to a rise in markups of 50 basis points, a contraction in employment of 3 percent and in output of 4 percent. These large effects suggest that government policies to extend financing to new entrants might have large effects on aggregate

Figure 28: Business applications in the Great Recession and 2020



employment. Such policies would increase hiring not only at new businesses but also at large incumbents, who, facing more competition from entrants, would lower markups and increase employment.

8 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 counter-cyclical 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.
- Baqae, 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.
- 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.
- Bloom, Nicholas.** 2009. "The Impact of Uncertainty Shocks." *Econometrica*, 77(3): 623–685.
- Clementi, Gian Luca, and Berardino Palazzo.** 2016. "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations." *American Economic Journal: Macroeconomics*, 8(3): 1–41.
- 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.
- 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. "In-

- vestment, 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.
- 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.
- 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.
- 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.
- 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 Compustat 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 (0.0008742***)	0.623711 (0.001559***)	0.375305 (0.001798***)
log XLR	0.885107 (0.003***)	0.688669 (0.005639***)	0.469273 (0.006349***)
log $COGS$	0.9164561 (0.0007804***)	0.780266 (0.001595***)	0.651581 (0.001949***)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry \times Year	Firm + Industry \times Year	Industry \times Year

Fact 2

Table 13: Variable input use and relative size over time

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

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 (0.0008417***)	0.632097 (0.001508***)	0.38278 (0.00174***)
log XLR	0.891063 (0.002387***)	0.683225 (0.005004***)	0.459426 (0.005529***)
log $COGS$	0.9334514 (0.0007165***)	0.79041 (0.00151***)	0.661271 (0.001869***)
Specification	Log levels	Log levels	Log difference
Fixed Effects	Industry \times Year	Firm + Industry \times Year	Industry \times Year

Fact 2

Table 15: Variable input use and relative size over time

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

B A simple model of entry and markups

In order to expose the key mechanism, in this section I study the Great Recession a simplified version of the quantitative model. I eliminate the firm life cycle and entry and exit, instead assuming that there is a fixed mass N_t of firms, each of which draws productivity from a fixed distribution $z \sim G$. This model differs in a few important ways from the quantitative model, but it is useful to explore the role of imperfect competition in amplifying the drop in employment during recessions.

Following [Edmond, Midrigan and Xu \(2018\)](#), there is a quick way to solve this model. The problem each firm faces can be rewritten in terms of a two random vari-

ables, one which summarizes aggregate and the other which summarizes idiosyncratic conditions. Solving the model then simply requires finding a fixed point in the aggregate variable. For more details, see Appendix E.1.

I calibrate the model in a simple procedure. First, as in the quantitative model, I assume that the productivity distribution is log-normal with mean parameter μ_z and dispersion σ_z . I choose the productivity distribution to be the stationary distribution of a log AR(1) with persistence parameter 0.5 and innovation variance 0.22. I choose the super elasticity parameter, ϵ/σ , so that the regression of log labor demand on log revenue matches the value in the data, of about 0.55, and I set σ so that the steady state markup with a unit mass of firms is 1.25. I run this regression on a sample that resembles compustat; that is, I use only the largest firms that comprise 30% of total labor demand. The parameterization is contained in Table 16.

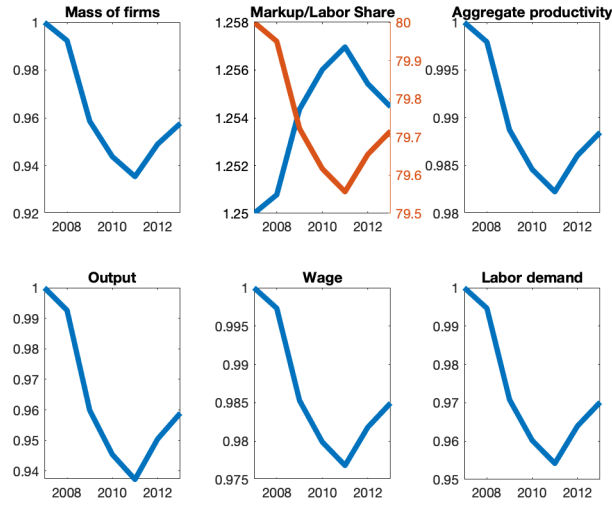
Parameter	Description	Value	Source
μ_z	Mean log productivity	0	Fixed
σ_z	Productivity dispersion	0.2540	Fixed
σ	Elasticity	7.1642	Calibrated
ϵ/σ	Super-elasticity	0.75	Calibrated

Table 16: Simple model parameterization

B.1 The Great Recession in the Model

The Great Recession saw a fall in the number of firms of about 6% and of establishments by around 4%. To study the effects of this change on employment and markups in this model, I feed in a path for the mass of firms N_t in the model to mimic the path of the mass of firms in the data. Figure 29 depicts the results of this exercise. The mass of firms falls gradually from 2007, eventually falling by over 6 percent. The bottom left panel shows that output falls by around the same amount, driven both by a fall in aggregate productivity of almost 2 percent and of labor demand by 5 percent. The wage falls by 2.5% - half as much as employment, owing to the GHH preference specification I use. The markup rises modestly, from 1.25 to nearly 1.259, which leads the labor share to fall from 80% to 79.4%. The rise in the markup means that the wage cannot fall enough to induce the remaining firms to maintain employment near its previous level.

Figure 29: The Great Recession in the simple model



Comparison to constant markups world

How much does variable markups contribute to business cycle dynamics? To understand the role of variable markups on employment, I compare the dynamics of this economy to a change in the mass of firms to one in which firms face a constant elasticity of demand. In the second economy, I use the CES version of the Kimball aggregator. To keep the models comparable, I calibrate the elasticity of substitution in the CES model so that the steady state cost-weighted markup is the same in each model.

Figure 30 depicts the results of this experiment. The labor share is constant in the CES version of the model. Aggregate productivity falls slightly more in the variable markups (“Kimball”) model. Labor demand falls by 2 percent more in the variable markups version of the model, accompanied by a steeper drop in output and the wage. In the CES version of the model, firms can be induced to hire more with a smaller drop in the wage. This then implies that employment does not fall by as much. However, with variable markups, firms respond to the fall in competition by cutting back on their employment, leading labor demand to fall by more. The wage cannot fall enough to induce them to hire as much as they would in the CES version of the model.

The quantitative importance of this mechanism appears to be large. Variable markups account for nearly one half of the 5% fall in labor demand caused by the fall in the number of firms. The fall in the number of firms leads to a significant fall in aggregate productivity, but variable markups do not seem to play a large role in that fall. This is consistent with [Arkolakis et al. \(2019\)](#), who show that the produc-

Figure 30: The role of variable markups



tivity effects of entry in a class of models that are similar to this one do not depend on whether markups are variable or fixed. Further note that this mechanism does not require large swings in the labor share - it falls by a modest 60 basis points.

It is worth discussing the role of household preferences. With a linear disutility of labor and log utility over consumption, the household's first order condition relates $W = \omega C$. As I show in Appendix ??, in this version of the model, employment does not fall at all in the CES version of the model in response to a shock to the mass of firms, but it does fall in the variable markups version. In the CES version of this model, the general equilibrium effect of wages exactly undoes the fall in labor demand from missing firms. With a constant Frisch elasticity of labor supply, I find results in between the GHH and linear disutility versions of the model. I show the GHH results here because they are best suited to a discussion of business cycles.

Increasing importance of the mechanism

In Section 3, I presented evidence that the markup-revenue relationship has grown stronger over time. This fact suggests that the mechanism I discuss here has become more important as market concentration has risen.

How would the Great Recession have been different had the relationship between markups and firm size been at its 1985 level? To answer this question, I change the value of the superelasticity to match the labor-revenue regression at the beginning of the Compustat sample. It turns out that this requires a superelasticity of $\epsilon/\sigma = 0.45$.

Figure 31: Shock to the number of entrants, 1985 vs 2015



Figure 31 compares the effects of the exogenous change in the mass of competing firms under these two parameterizations. The labor share falls in both parameterizations, though the initial values of the labor share are different. In the 2015 calibration, it falls from 80% to 79.4%, and in the 1985 calibration, it falls from 81.7% to 81.3%. Aggregate productivity in each falls by around the same amount. Labor demand falls by an extra 1.5% more in the 2015 calibration, and wages and output also fall by more.

C Robustness in the simple model

C.1 The choice of superelasticity

Figure 32 depicts the markup and aggregate productivity at different values of the number of firms. As the left panel shows, the markup falls as the number of firms increases, and, as the right panel shows, aggregate productivity rises with the number of firms. These changes are quantitatively significant; a 6 percent fall in the mass of firms leads the markup to rise by 87 basis points and productivity to fall by 1.5%. The rise in the markup implies a fall in the labor share of 55 basis points.

How sensitive is this number to the super-elasticity parameter? There is considerable heterogeneity across the literature in the magnitude of the superelasticity of demand. As Table 17 shows, estimates in the literature vary from 0.14 to 0.6. So, next, I re-do this exercise, varying the super-elasticity and re-calibrating the average elasticity for each value of the super-elasticity to match a steady state markup of 1.25.

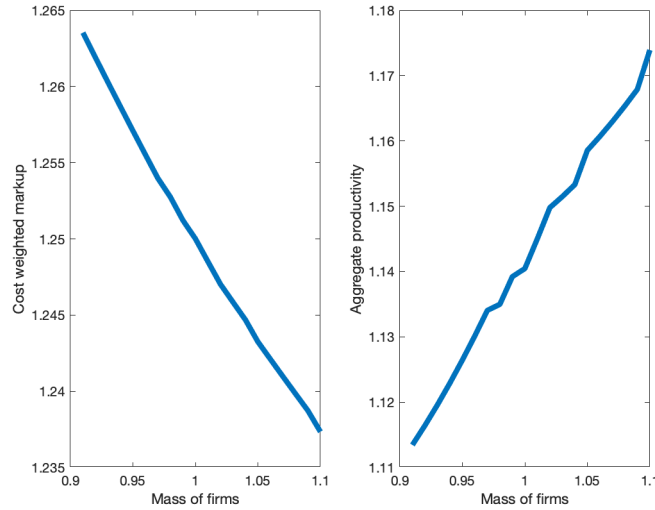


Figure 32: Mass of firms, the markup, and productivity

Paper	σ	ϵ	ϵ/σ
This paper			
Edmond, Midrigan and Xu (2018)	10.18	1.4252	0.14
Amiti, Itskhoki and Konings (2019)	5	1.6	0.26
Berger and Vavra (2019)	5	2.35	0.47
Gopinath, Itskhoki and Rigobon (2010)	5	3	.6

Table 17: Selected parameterizations of [Klenow and Willis \(2016\)](#) demand

Figure 33 depicts the results. As this figure shows, the super-elasticity affects the degree to which the mass of operating firms affects the markup but not the value of productivity. For the markup, the effect seems non-linear; there is a large difference between a super-elasticity of 0.15 and 0.5, but not much of a difference between 0.5 and 1. The range of markups at a mass of firms equal to 0.94 is 1.253 to 1.267. Productivity varies strongly with the number of firms and does not appear to depend much on the size of the super-elasticity. This is consistent with the findings of [Arkolakis et al. \(2019\)](#), who find that the productivity effects of increasing competition do not differ much between a [Klenow and Willis \(2016\)](#)-like world and one with CES demand.

In Figure 29, I depict the results of a simple experiment: I exogenously change the number of competing firms. The left panel shows the path of the number of firms, starting in 2007. This is the number of firms operating in the BDS, relative to its value in 2007. As the middle panel shows, the cost-weighted markup rises by 1 percentage point, which implies a fall in the labor share from 80% to 79.4%. The right panel shows

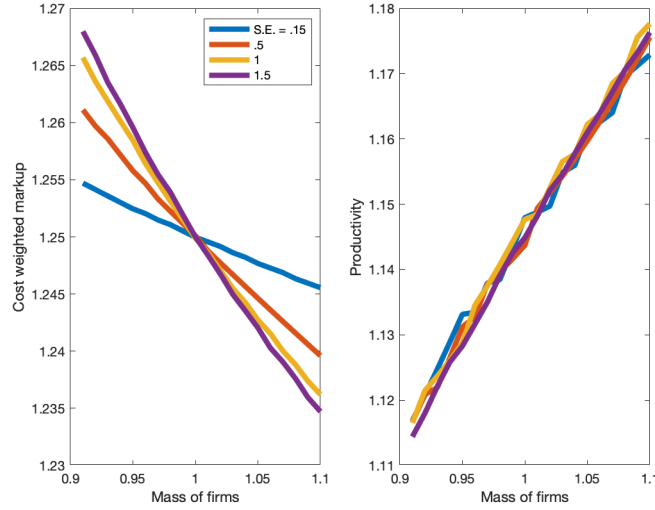


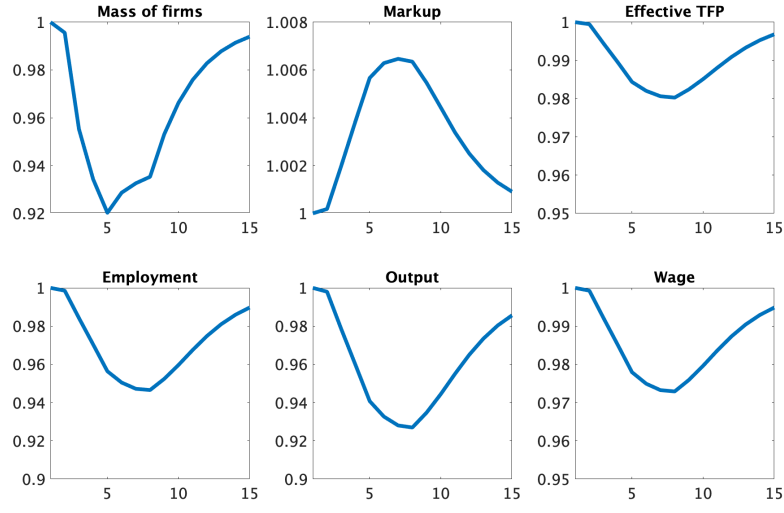
Figure 33: Mass of firms, the markup, and productivity for different values of the superelasticity

that aggregate productivity falls by 1.8 percent. Importantly, labor demand falls by 5 percentage points over 3 years. Since the number of firms operating during the Great Recession fell persistently and took a long time to recover, these changes peak in 2011 and remain substantially different from their 2007 levels through 2013.

D 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 34: The Great Recession shock to the entry of firms



E Solution method

E.1 Simple model

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

$$\pi(z) = \max_{y \geq 0} \left[\Upsilon'(y/Y) Dy - W \frac{y}{z} \right] \quad (\text{E.1})$$

$$= \max_{y \geq 0} \left[\Upsilon'(y/Y) Dy / YY - W \frac{y / YY}{z} \right] \quad (\text{E.2})$$

$$= \max_{q \geq 0} \left[\Upsilon'(q) DqY - W \frac{qY}{z} \right] \quad (\text{E.3})$$

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

$$(\text{E.5})$$

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

$$\tilde{\pi}(z) = \max_{q \geq 0} \left[\Upsilon'(q) q - Aq \right] \quad (\text{E.6})$$

(E.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.

E.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 .

F Pareto vs. Log-normal

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

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

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

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

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

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

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

Performing the change of variables implies that:

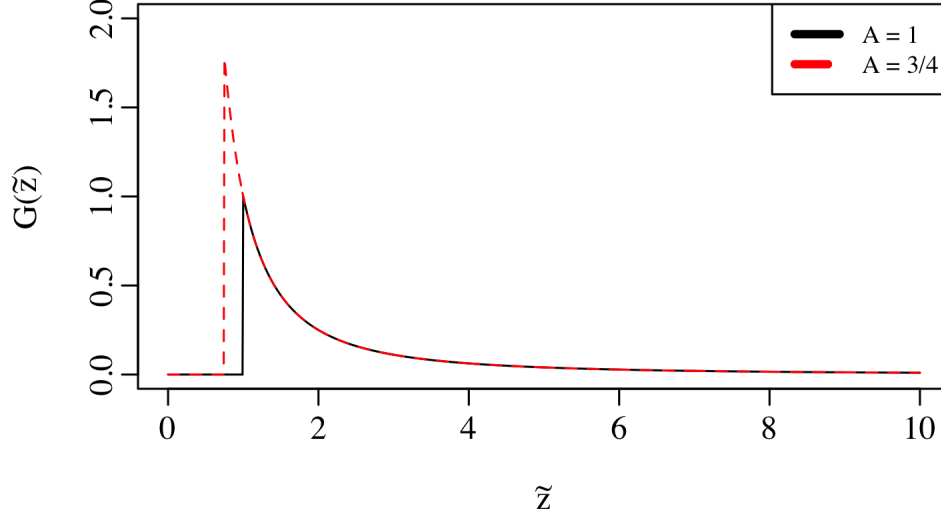
$$G(\tilde{z}; \eta, \theta) = 1 - \left(\frac{\eta}{Az} \right)^\theta \tag{F.1}$$

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

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

A change in A thus only affects the location of the Pareto distribution (up to rescaling). I show an example of this kind of shift in Figure 35

Figure 35: A change of variables under the Pareto assumption



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?

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

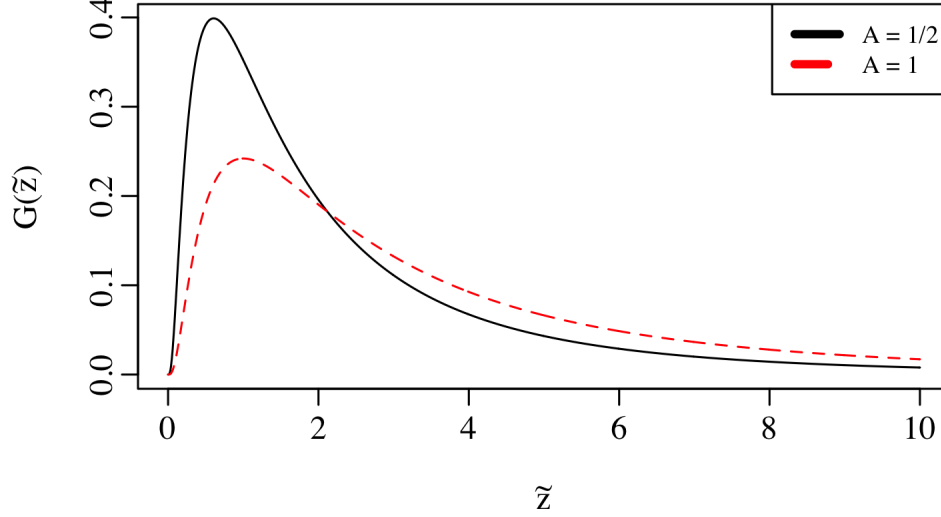
Suppose that $\log z \sim \mathcal{N}(\mu, \sigma^2)$. A change of variables implies that $\log \tilde{z} \equiv \log Az \sim \mathcal{N}(\log A + \mu, \sigma^2)$.

Recall the variance of a log-normally distributed variable:

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

An increase in $\log A$ then increases both the mean and variance of \tilde{z} . Figure 36 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

Figure 36: A change of variables under the log-normal assumption



in the markup.

G 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.

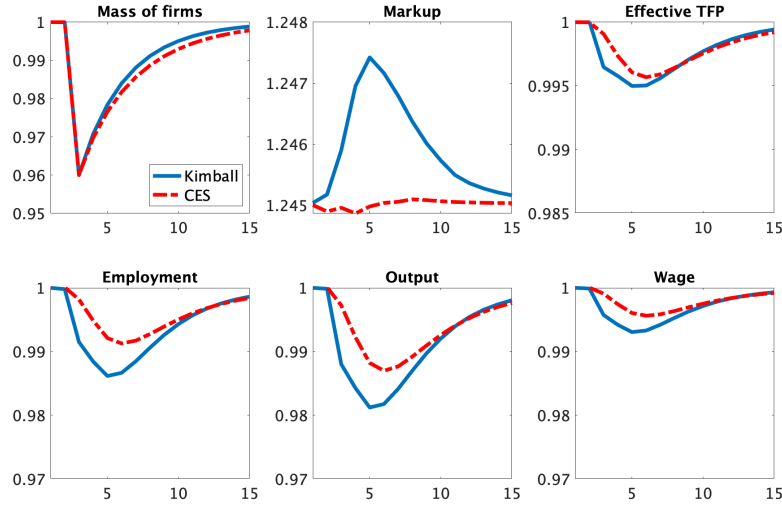
Figure 37 summarizes the effects of a shock to the cost of entry in the Kimball and CES models. The figure looks qualitatively similar to the behavior of aggregates following a shock to the mass of potential entrants.

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

$$V_E(\phi) \geq c_E$$

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 $\phi \geq \tilde{\phi}$.

Figure 37: Shock to the cost of entry



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.

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 38 depicts the paths of these variables.

The entry mass shock, on the other hand, reduces the average size of entrants as well. Figure 39 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 38: Entrants following the entry cost shock

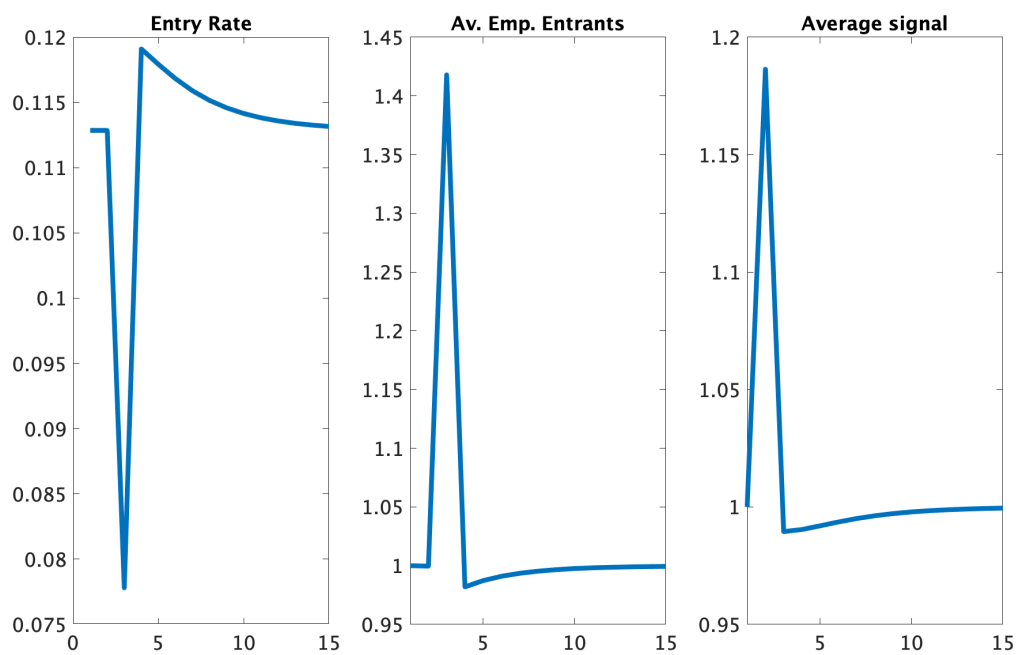
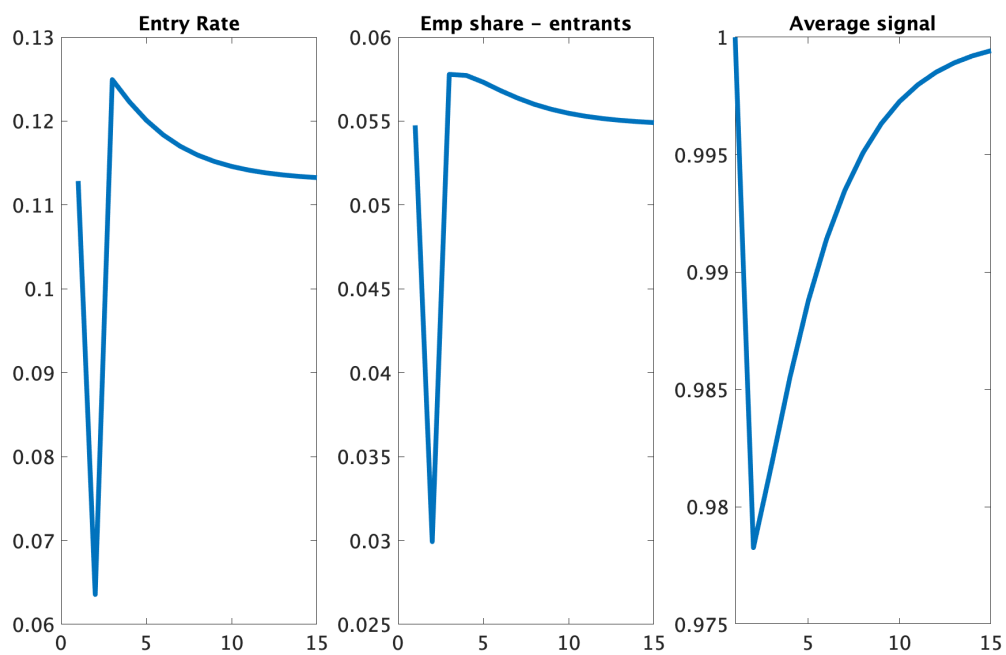


Figure 39: Entrants following the entry mass shock



H 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.

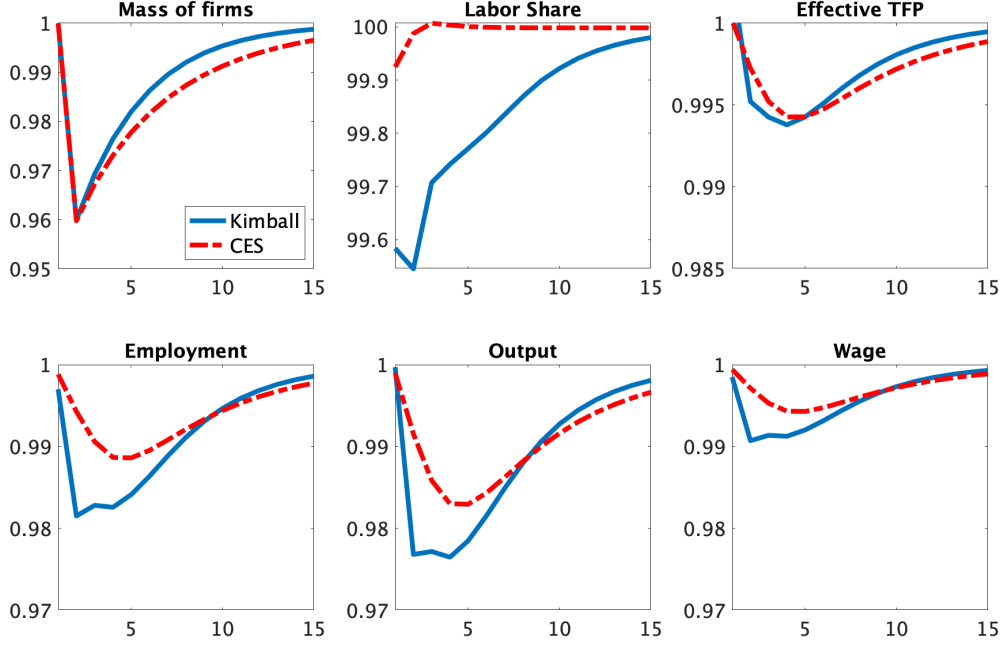
Let φ denote the cost of issuing equity (i.e., of having negative profits). Moreover, suppose that there is a deterministic component to the fixed cost \bar{c}_F ¹¹. The firms' flow value is:

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

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 H 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.

¹¹There is still some component of the fixed cost that is random. This is useful for targeting the average size of exiting firms.

Figure 40: Financial shock



I Stochastic Discount Factor

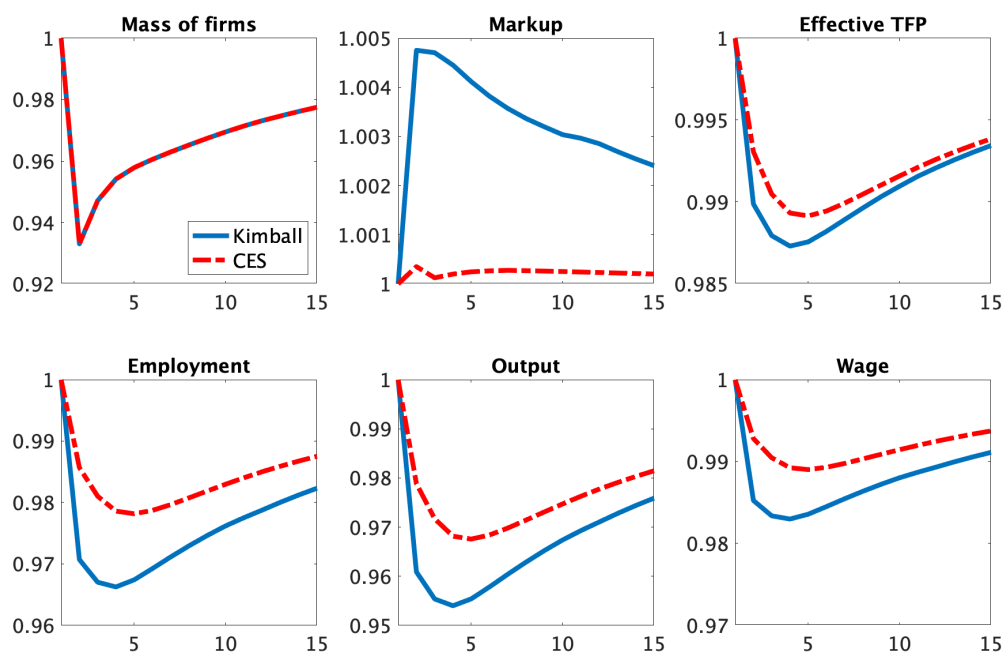
I.1 Shock to entry

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 41. 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 41: Impulse response to an entry shock; variable stochastic discount factor



I.2 TFP shock

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

