The Importance of Entry in the Business Cycle: What are the Roles of Markups, Adjustment Costs, and Heterogeneity?

William Gamber

Federal Reserve Board

June 7, 2023

Abstract

In this paper, I evaluate the role of fluctuations in business formation in amplifying business cycles. To do this, I study the response of employment to shocks in a general equilibrium model of producer dynamics with entry and exit. The model features producer heterogeneity, adjustment frictions, and a variable demand elasticity. I find that, while heterogeneity reduces the effects of entry on the broader economy, the variable demand elasticity and adjustment frictions amplify these effects, so that entry fluctuations lead to economically meaningful amplification of business cycle shocks.

1 Introduction

During the Great Recession, the number of new businesses created each year declined more than 35 percent relative to its peak in the mid 2000s and remained depressed through 2019, leading to a large and persistent decline in the number of operating businesses.[1] This fall in entry accompanied a decline in employment per capita of over 8 percent that only slowly returned to its pre-recession level. In this context, it is natural to ask: what is the role of entry in amplifying business cycles?

In this paper, I answer this question using a quantitative general equilibrium model of producer dynamics. The model is consistent with three important features of the data. First, the model features heterogeneous producers, reproducing the life-cycle pattern of producer size and features of the cross-sectional distribution of revenue that we observe in US data. Second, in the model, producers face an elasticity of demand that declines with relative size, replicating the relationship between revenue and variable cost use that we observe in data. And third, producers in the model face adjustment frictions that slow the reallocation of inputs, consistent with producer-level employment dynamics in the data.

In the model, a shock to the cost of entry that leads business formation to fall as much as it did in the United States during the Great Recession generates a decline in employment of nearly 2.5 percent. I also quantify the extent to which variations in entry amplify TFP shocks in the model, finding that an exogenous decline in TFP leads to a decline in aggregate employment of about 50 percent more in this model relative to a model with no entry.

I then study the roles that producer heterogeneity, adjustment frictions, and the

I am very grateful to the editor, Greg Kaplan, and two anonymous referees for their comments. I am also very grateful to my advisors Simon Gilchrist, Ricardo Lagos, and Virgiliu Midrigan and committee members Corina Boar and Mark Gertler for their guidance and support throughout this project, which began as a part of my dissertation. I would also like to thank Ryan Decker, Sebastian Graves, James Graham, Dino Palazzo, Michael Siemer, Melinda Suveg (discussant), and participants at several seminars and conferences for their comments. The views expressed in this paper are solely those of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System. This paper was edited by Greg Kaplan.

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

variable elasticity of demand all play in determining the size of these effects. Producer heterogeneity reduces the effect of entry on aggregate employment because entering producers are significantly smaller than incumbents. Adjustment frictions, on the other hand, amplify the effect of the decline in entry on employment; these costs prevent incumbent producers from expanding following the decline in the wage caused by the fall in entry. And lastly, the variable elasticity of demand, which leads incumbent producers' markups to rise and labor demand to fall after the decline in entry, plays a smaller role in amplifying the decline in employment.

This paper contributes to an existing literature that uses structural models to study the role of entry in business cycles. Papers in that literature come to very different conclusions about the importance of entry depending on their modeling assumptions. Papers that study the role of entry in models in which firms are homogeneous typically find large effects of entry on business cycle fluctuations. (Bilbiie, Ghironi, and Melitz (2012) and Jaimovich and Floetotto (2008)) More recent work studies entry fluctuations in theoretical models with a realistic firm lifecycle but without a variable elasticity of demand or labor adjustment frictions. (Clementi and Palazzo (2016) and Lee and Mukoyama (2018)) These papers find a much more modest role for entry. Lastly, Edmond, Midrigan, and Xu (2018) study entry in a model with heterogeneous firms and variable markups but no variable input adjustment costs and find almost no effect of entry on the aggregate markup.

This paper provides a framework for understanding these disparate results. A version of the model without variable input adjustment costs or a variable demand elasticity, similar to Clementi and Palazzo (2016), implies effects of entry on employment that are about two-thirds smaller than in the baseline model on impact. A version of the model without heterogeneity, similar to Bilbiie, Ghironi, and Melitz (2012), implies effects of entry on employment that are up to twice as large as in the baseline model. Lastly, a version of the model with a variable elasticity of demand but no adjustment costs, similar to Edmond, Midrigan, and Xu (2018), generates fluctuations in employment following a decline in entry that are less than half the baseline model on impact.

I begin the paper by describing the model. The model is a general equilibrium Hopenhayn (1992) model with several features, including (1) a variable elasticity of demand, (2) labor adjustment costs, (3) a producer lifecycle, and (4) heterogeneity in size among producers, even after conditioning on age. Producers in the model have ex-ante heterogeneous, stochastic productivity. They are each the monopolistic supplier of a differentiated variety and face downward sloping demand with an elasticity that declines with relative size. The shape of the demand curve implies that producers have an incentive to increase their markup when their output rises relative to the overall market. Producers must pay a convex net hiring and firing cost, which slows their responses to idiosyncratic shocks and prevents inputs from rapidly reallocating across businesses. Lastly, businesses exit each period and are replaced, in steady state, by newly created businesses.

I then parameterize the model. One important parameter is the superelasticity of demand, which governs how quickly the elasticity of demand falls with producer size. My approach to quantifying that parameter is motivated by the "production function approach" (PFA) that has been popular in the recent macroeconomics literature on markups (see De Loecker, Eeckhout, and Unger (2020), for example). The intuition behind this approach is that, under the assumption that producers can frictionlessly adjust their variable inputs, the wedge between variable input use and revenue is informative about the size of the markup. I show that this wedge varies with producer size in the data; the typical producer in the sample increases its variable input bill much less than one-for-one with its sales. This finding suggests that markups rise with producer size.

I use the quantified model to structurally interpret these regressions. As highlighted by Bond et al. (2021), the PFA requires the restrictive assumption that variable inputs can be costlessly adjusted. To relax this assumption, I use the reduced-form regression coefficients, along with data on employment reallocation, to discipline parameters in the model, including the degree of adjustment costs and the extent to which the elasticity of demand falls with producer size. I then simulate a panel of producers in the model and estimate the same PFA regressions on simulated data. I show that not accounting for adjustment costs leads to an overstatement of the relationship between producer size and markups but that large producers' markups do vary significantly with their size, with an elasticity of the markup to relative sales of around 30 percent.

To quantify the effects of fluctuations in entry on aggregate employment, I then introduce a shock to the cost of entry to the model. This shock leads to a temporary decline in entry that has economically meaningful and persistent effects on aggregate employment. The fall in entry increases the market shares of incumbent producers and leads them to increase their markups, produce less, and reduce employment. The decline in entry also leads the wage to fall, which induces incumbent producers to increase their labor demand. Labor adjustment costs, however, prevent them from hiring quickly. Love for variety effects mean that the decline in the number of operating producers reduces aggregate productivity. The movements in the markup and productivity are economically significant; in response to a shock that reduces entry by one-third, as much as it fell during the Great Recession, the aggregate markup rises 0.9 percent and aggregate productivity falls 0.6 percent. Because of these changes, employment declines about 2.5 percent.

I next study the mechanisms in the model that generate these large fluctuations in employment in response to the fall in entry. I show that adjustment costs and the variable elasticity jointly account for most of this response – a frictionless model with a constant elasticity has a 70 percent smaller employment response to a shock to entry. The difference between that model and the baseline model arises primarily because of fluctuations in the aggregate markup.

Adjustment costs and the variable elasticity explain about 50 percent and 20 percent of the overall employment response, respectively. The variable elasticity amplifies the entry cost shock because, as entry falls and incumbent producers' relative output rises, incumbents increase their markups and reduce their labor demand. This producer-level increase in markups is somewhat offset by a reallocation of employment away from high-markup and to low-markup producers. Adjustment costs, on the other hand, amplify the propagation of the entry cost shock in two ways. First, as the wage falls in response to a decline in entry, they prevent producers from immediately hiring more workers. And second, adjustment costs amplify the effect of the variable elasticity by slowing the reallocation from high-markup to low-markup producers.

To understand the role of heterogeneity in the propagation of entry fluctuations to employment, I compare the baseline model to two alternative models that omit key aspects of producer heterogeneity. In the first alternative model, I assume there is no producer lifecycle: entering producers are the same size as incumbents, on average. TFP moves much more in response to the shock in that economy, leading the effect on aggregate employment to be roughly double in that economy relative to the baseline. I then compare the baseline to models with no heterogeneity, as in Bilbiie, Ghironi, and Melitz (2012), finding that a shock to entry in those models has effects on employment that that are 1.5 to 2 times the size of those in the baseline economy.

Having studied the transmission of entry fluctuations to aggregate employment in isolation, I then study the response of the economy to an exogenous shock to TFP that leads to endogenous movements in entry. The decline in TFP leads entry to fall, the markup to rise and employment to fall. I show that endogenous entry fluctuations lead aggregate employment to fall by 50 percent more relative to a no-entry baseline, driven largely by a rise in the markup. I thus find that entry plays an economically meaningful role in business cycles.

2 Quantitative Model

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

2.1 Environment

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

2.2 Household

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

$$\sum_{t=0}^{\infty} \beta^t u(C_t, L_t) \tag{1}$$

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

$$C_t \leq W_t L_t + \Pi_t. \tag{2}$$

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

$$W_t = -\frac{u_{L,t}}{u_{C,t}}. \tag{3}$$

2.3 Final goods producer

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

$$\int_0^{N_t} \Upsilon\left(\frac{y_t(\omega)}{Y_t}\right) d\omega = 1, \tag{4}$$

where $\Upsilon(q)$ is a function that satisfies three conditions: it is increasing ($\Upsilon'(q) > 0$) and concave ($\Upsilon''(q) < 0$), and $\Upsilon(1) = 1$. Given quantities of each intermediate variety $\{y_t(\omega)\}$, aggregate output Y_t is defined as the solution to Equation (4).

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

$$p_t(\omega) = \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right) D_t, \tag{5}$$

where the D_t is the demand index, defined as

$$D_t \equiv \left(\int_0^{N_t} \Upsilon'\left(\frac{y_t(\omega)}{Y_t}\right) \frac{y_t(\omega)}{Y_t} d\omega\right)^{-1}. \tag{6}$$

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

$$\Upsilon(q) = 1 + (\sigma - 1) \exp\left(\frac{1}{\epsilon}\right) \epsilon^{\frac{\sigma}{\epsilon} - 1} \left[\Gamma\left(\frac{\sigma}{\epsilon}, \frac{1}{\epsilon}\right) - \Gamma\left(\frac{\sigma}{\epsilon}, \frac{q^{\epsilon/\sigma}}{\epsilon}\right)\right] \tag{7}$$

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

$$\Gamma(s, x) = \int_x^{\infty} t^{s-1} \epsilon^{-t} dt. \tag{8}$$

This specification of Υ generates an elasticity of demand for each variety that is decreasing in its relative quantity $q_t \equiv y_t / Y_t$ so that large producers set higher markups than small producers.[2] Under the Klenow and Willis (2016) specification,

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

$$\Upsilon'(q) = \frac{\sigma - 1}{\sigma} \exp\left(\frac{1 - q^{\frac{\epsilon}{\sigma}}}{\epsilon}\right) \quad (9)$$

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

2.4 Intermediate goods producers

At each date t, there is a mass N_t of intermediate goods producers, each of whom is the monopolistic supplier of a differentiated variety ω. Each hires labor in a perfectly competitive market at wage W_t, produces its variety using a constant returns production function, and sells it to the final goods producer, taking as given the demand schedule.

Timing works as follows: in each period, each producer observes its idiosyncratic productivity z_t and the state of the aggregate economy, Λ_t. It then hires workers, produces output, and sells its differentiated variety to the final goods producer. Producers face labor adjustment costs $\phi(\ell, \ell')$ as a function of last period's employment ℓ and their current employment ℓ'. After selling their output and paying adjustment costs, each producer draws an i.i.d. exit shock; with probability $\mathbb{P}(exit)$, the producer is forced to exit. The value of exit is normalized to 0. They discount future streams of profits using the discount factor m.[3]

[3] In the deterministic steady state, the producer discounts future steams of profit at rate β, regardless of the household's stochastic discount factor. Later in the paper, I study deterministic dynamics. For my baseline results, I assume that producers discount future streams of profits using the risk neutral discount factor β. This assumption is equivalent to assuming either (1) the economy is small and open so its interest rate is fixed or (2) all producers are owned by a measure zero, risk-neutral mutual fund that distributes profits to households. The reason that I choose a risk-neutral discount rate is that the preference specification I use counterfactually implies that interest rates rise in recessions. As emphasized in Winberry (2021), interest rates are pro–cyclical, consistent with a countercyclical discount factor. In this paper, as in Winberry (2021), the interest rate affects producer dynamics. To avoid mischaracterizing the effect of falling entry on aggregate employment, I fix the discount rate and thus the interest rate.
In appendix G, I study the response of the economy to aggregate shocks when producers price streams of profit using the household's stochastic discount factor. In response to the decline in entry, consumption initially falls and returns to its steady state. Under the household preferences that I use, this movement leads the discount factor to fall. The decline in the discount factor has two effects that amplify the response of the economy to entry shocks: (1) it decreases the value of entry further and thus deepens and prolongs

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

$$V(\ell, z; \Lambda) = \max_{p,\ell'} \pi(z, \ell', p; \Lambda) - \phi(\ell, \ell') + \beta \mathbb{P}(exit)\mathbb{E}\Big[m' V(\ell, z'; \Lambda) \big| z\Big], \tag{10}$$

$$\pi(z, \ell', p; \Lambda) = \left(p - \frac{W}{\ell'}\right) d(p; \Lambda), \tag{11}$$

$$y \leq z\ell'. \tag{12}$$

Equation (10) shows that the value of a producer is its period profits $\pi(\cdot)$, less the adjustment costs it pays $\phi(\cdot)$ and plus its continuation value. The continuation value equals the expected value of operating next period, discounted using the household's stochastic discount factor and the exogenous producer destruction rate. Equations (11) and (12) describe the production function and demand system that producers face.

2.5 Entrants

There is free entry in the model. Each period, an unlimited mass of potential entrants considers whether to begin producing. Each potential entrant observes the aggregate state of the economy and decides whether to pay a sunk cost $c_E(\Lambda_t)$ to enter.[4] After paying the sunk cost, each entrant draws a value for idiosyncratic productivity from a distribution $H(z)$, freely hires labor, and immediately produces and sells output.[5]

The value of an entrant who has paid the sunk entry cost is:

the fall in entry and (2) it makes producers more hesitant to hire.

[4] Note that this sunk cost could vary arbitrarily with the aggregate state of the economy. Later in the paper, I will impose a functional form for $c_E(\Lambda_t)$.

[5] An alternative model of entry would be the selection model presented in Clementi and Palazzo (2016). In that model, each potential entrant observes a signal of its productivity after entry and then decides whether to enter. As I show in Appendix C, this alternative has two important counterfactual implications: (1) the entry rate exhibits significantly less volatility than it does in the data and (2) the share of employment among entrants and young firms varies too little relative to the US data.

$$V_E = \int_z \max_\ell V(\ell, z) dH(z). \tag{13}$$

The optimal policy of the potential entrant is to enter if and only if $c_E(\Lambda_t) \leq V_E$. In equilibrium, potential producers will enter until the sunk cost of entry equals the value of entry – at which point potential entrants are indifferent between entering and not.

2.6 Aggregation

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

$$Y_t = Z_t L_t. \tag{14}$$

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

$$Z_t = \left(\int \int \frac{y_t(z, \ell)}{Yz} d\Lambda_t(z, \ell) \right)^{-1}. \tag{15}$$

This quantity grows with the number of producers (love for variety) and with the extent to which output is produced primarily by high–productivity producers. The aggregate markup is implicitly defined as the inverse labor share:

$$\mathcal{M}_t = \frac{Y_t}{W_t L_t}. \tag{16}$$

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

$$\mathcal{M}_t = \int \int \mu_t(z, L) \frac{\ell_t(z, L)}{L_t} d\Lambda_t(z, L). \tag{17}$$

3 Steady State

In this section, I discuss how I parameterize the model in order to match key features of the cross-section distribution of producers as well as within-producer dynamics. I also explore the steady state properties of the model, including lifecycle dynamics.

3.1 Markups and producer size

One feature of the model is that the demand elasticity faced by producers declines with their relative output, so that markups rise with size. To quantify this mechanism, I use an empirical strategy motivated by the production function approach (PFA). (See, for example: De Loecker and Warzynski (2012) and De Loecker, Eeckhout, and Unger (2020))

Consider a producer, p, in an industry i, at date t, with a production function in potentially many inputs. Denote by α_{ipt} the elasticity of output with respect one of its inputs X, which has price P^X. There are two important assumptions underlying the approach:

Assumption 1 *Producers' output elasticities are equal to the product of a permanent producer component (α_p) and a time-varying industry component (α_{it}); that is, $\alpha_{ipt} = \alpha_p \times \alpha_{it}$.*

Assumption 2 *The producer faces no adjustment cost on its variable input X.*

Then, a first-order condition of the cost minimization problem with respect to X gives a relationship between the total amount spent on that variable input $P^X X$, revenue PY, the markup μ, and the output elasticity:

$$(P^X X)_{ipt} = \alpha_{ipt} \frac{(PY)_{ipt}}{\mu_{ipt}}. \tag{18}$$

Taking logs of this first order condition gives the following equation:

$$\log(P^X X)_{ipt} = \log \alpha_p + \log \alpha_{it} + \log(PY)_{ipt} - \log \mu_{ipt}. \tag{19}$$

Equation 19 shows that a larger covariance between (log) markups and (log) revenues at the producer level generates a weaker relationship between revenue and total variable cost. Differencing within producers over time then motivates the following specification:

$$g((P^X X)_{ipt}) = \tilde{\alpha}_{it} + \beta \times g((PY)_{ipt}) + \epsilon_{ipt}, \tag{20}$$

where $g(\cdot)$ denotes growth rate, $\tilde{\alpha}_{it}$ is an industry-time fixed effect and ϵ_{ipt} is a residual.

Data. I use a panel of publicly listed, US-based firms in Compustat. I restrict the sample to observations between 1985 and 2018, exclude financial firms and utilities, and for my baseline results classify firms using the Fama-French-49 industry definitions.[6]

This sample, while not representative of the average firm in the economy, accounts for a large portion of US output and employment. Firms in this sample are only 1 percent of firms in the United States, but their sales equal roughly 75 percent of nominal gross national income and their total employment accounts for 30 percent of nonfarm payrolls. In my baseline results, I use the cost-of-goods-sold (COGS) as a measure of variable input costs. COGS includes materials and intermediate inputs, labor costs, energy, and other expenses associated with the production of the firm.

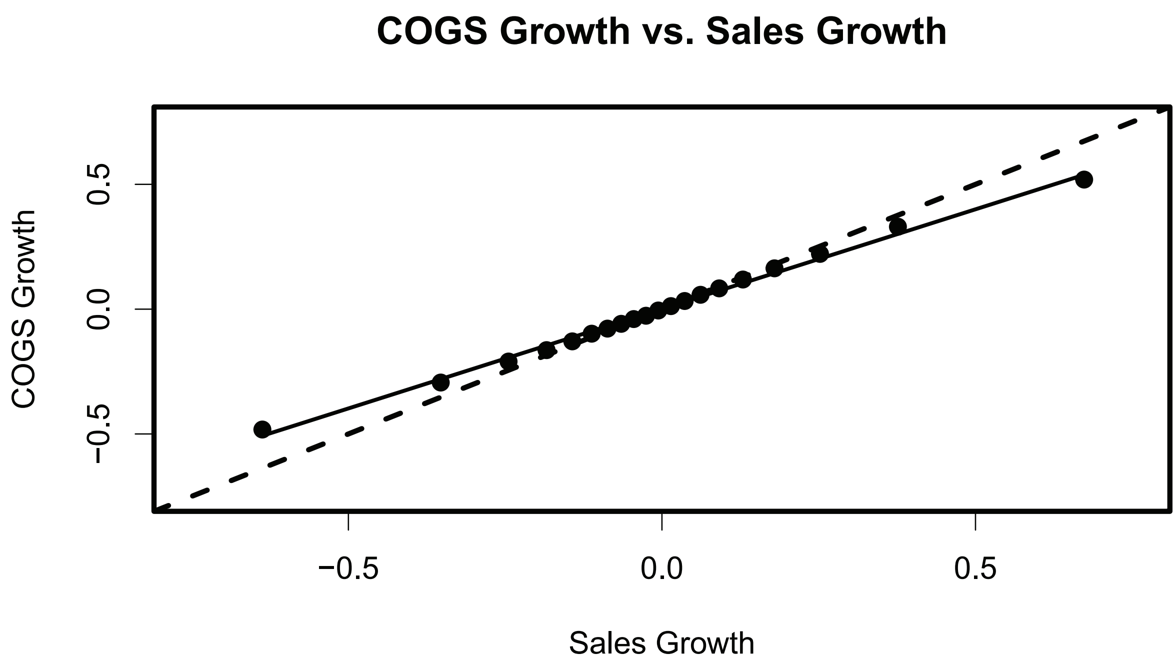
Results. I estimate $\hat{\beta} = 0.654$ (0.002).[7,8] Figure 1 depicts a binscatter plot of this

[6] The results that follow are not sensitive to the definition of industry – in Appendix A, I show that similar results hold using SIC and NAICS definitions at various levels of granularity.

[7] Standard error in parentheses. Reported results use growth rates defined as $g(x_{ipt}) = 2(x_{ipt} - x_{ipt-1})/(x_{ipt} + x_{ipt-1})$, as is common in the literature on firm dynamics; see for example: Davis, Haltiwanger, and Schuh (1996). This specification is similar to log growth rates but is bounded between -2 and 2 and so reduces the influence of outliers. See Appendix A for more details on the regression, as well as estimates for a variety of specifications for the variable cost and choices of fixed effects.

[8] In contrast to De Loecker, Eeckhout, and Unger (2020), I do not estimate the elasticity of output with respect to COGS, instead allowing fixed effects to pick up variation in α across firms and over time. This avoids needing to compute a measure of real output for each firm, which, as Bond et al. (2021) point out, is not feasible in Compustat, where we only observe revenue.

Figure 1: COGS and Sales Growth in Compustat



Note: The figure a binscatter of COGS and revenue growth, residualized by industry time fixed effects. The solid line depicts the fitted values from estimating Equation 20. The constant markup benchmark (dashed line) is a 45-degree line. Source: author's calculations

regression. Under assumptions 1 and 2, the estimated coefficient of 0.654 is consistent with an elasticity of the markup to firm size of 34.6 percent. If markups did not vary at all with firm size, then the regression coefficient would be 1, represented by the dashed line in figure 1. The data are consistent with a lower coefficient, shown by the solid line.

The output elasticity assumption. A key assumption in this framework is that the output elasticity is the product of a permanent producer component and an industry-time component. Under this assumption, fixed effects absorb the output elasticities in the regression I estimate.

However, if producers' output elasticities vary over time relative to the industry-year mean, then part of the deviation of the regression coefficient from 1 could reflect variation in the output elasticity, rather than the markup. For example, consider a producer whose markups are constant and who faces no adjustment frictions. If its output elasticity declines with relative size, then its variable cost bill will grow by less than one-for-one with revenue. So, the estimated coefficient in equation 20 could potentially reflect variation in the output elasticity rather than the markup.

Assumption 1 is similar to those standard in the literature and is difficult to relax without directly estimating the output elasticity. Estimating the output elasticity requires observations of output, not revenue, as discussed in Bond et al. (2021), and balance sheet data like Compustat do not provide a measure of output.

The frictionless assumption. Assumption 2, that the variable input is costlessly adjustable, is violated in the model I study. Moreover, adjustment costs could affect the estimated coefficient in equation 20. For example, consider a firm with an infinite labor adjustment cost. In response to an increase in productivity, the firm would increase its revenue without changing its employment at all, implying a regression coefficient of 0. Under assumption 2, one would mistakenly conclude that this firm increases its markups one-for-one with its relative size.

For a firm facing adjustment costs on its variable input (that is, if that input is not truly variable), the static first order condition in the PFA does not hold. In that case, the quantity μ would represent any wedge distorting the firm's production choices away from its static optimum – not just the markup. To avoid misattributing variation in this wedge to variation in the markup, I will use the model to discipline my interpretation of these regression coefficients. When I calibrate the model, I will jointly choose both the superelasticity of demand and the magnitude of adjustment costs to match both the estimated coefficient in this regression and additional data on firm–level labor adjustment dynamics. This strategy allows me to interpret these regressions in a structural model with adjustment costs.

Markups and producer size in the model. In the model, the elasticity of demand falls with relative size, so that producers have an incentive to increase their markups as they grow relative to the market. To measure the strength of this mechanism, I choose parameters in the model to match the estimate from equation 20.

To understand the role of this mechanism, whose strength is dictated by the superelasticity of demand ϵ/σ, consider the model without adjustment costs. In that case, each producer's only idiosyncratic state variable is its productivity. As a producer's productivity rises, it produces more and its elasticity of demand falls. In response, it increases its markups. The increase in markups means that the producer increases its employment less than one-for-one with its sales.

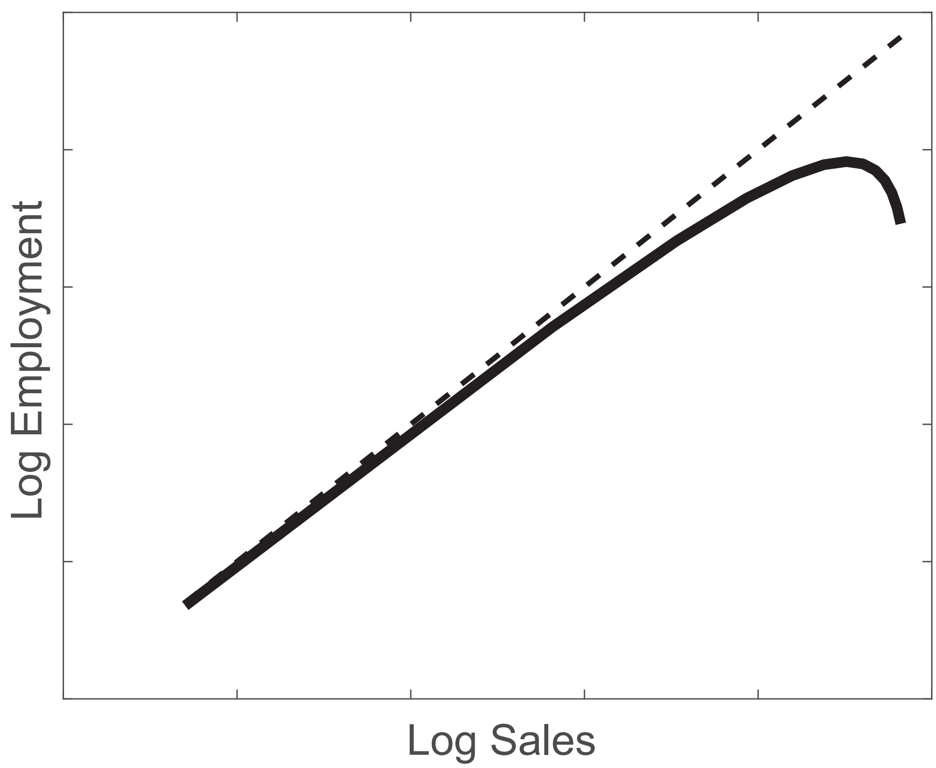
Figure 2 depicts the relationship between sales and employment in this model in blue and the same relationship in a model with constant markups in the black dashed line. Producers in the variable markup model increase their markups as their sales grow, which implies that the slope of the sales-employment relationship is always less than one. Because larger producers increase their markups more with sales than small producers do, this relationship is also concave. For the largest producers, markups increase so much with sales that their employment actually falls as they gain market share.

While the relationship between sales and employment is evidently non-linear in the model, I target the estimate of a linear regression of variable input growth on sales growth in the data. This discrepancy presents a challenge in calibrating the model, as the average Compustat producer is larger than the average producer in the economy, which might lead me to overstate the extent to which markups rise with market share for the average producer. To calibrate the model, I therefore choose parameters so that equation (20) estimated on a sample of the 1% largest producers in the model matches the regression from data. This procedure generates a comparable subsample to estimate the super-elasticity.[9]

Adjustment costs also affect this regression coefficient: a higher adjustment cost leads input use to vary less with revenue. To identify the size of adjustment frictions, I also require the model to match the autocorrelation of within-firm employment growth.

[9] In the baseline model, these producers account for 26% of sales.

Figure 2: Employment and sales in the frictionless model



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

A higher value of the adjustment cost leads producers in the model to gradually respond to idiosyncratic shocks, increasing the autocorrelation.

3.2 Calibration targets and results

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

$$u(C_t, L_t) = \frac{1}{1-\gamma}\left(C_t - \psi\frac{L_t^{1+\nu}}{1+\nu}\right)^{1-\gamma}. \tag{21}$$

I also impose a quadratic form for the labor adjustment cost, with a fixed employment depreciation rate δ. The size of the adjustment cost is determined by the parameter ϕ_ℓ.

$$\phi(L, L_{-1}) = \phi_\ell\left(\frac{L-(1-\delta)L_{-1}}{(1-\delta)L_{-1}}\right)^2 L_{-1}. \tag{22}$$

I assume that the natural logarithm of productivity follows an AR(1) process with persistence ρ_z and innovation variance σ_z^2. Entering producers draw their initial productivity value from a shifted version of the stationary distribution implied by the law of motion for incumbent productivity, $G(\log(z))$. In particular, entering producers draw their initial value of log productivity from the distribution $H(log(z)) = G(log(z)+d_E)$. I choose the parameter d_E to match the average employment of entering establishments relative to the overall average in the BDS.[10]

Calibration strategy. I fix five parameters and then jointly choose the remaining parameters to ensure that the model is consistent with salient features of the data. The pre-set parameter choices are summarized in table 1. I then simultaneously choose the productivity innovation persistence ρ_z and dispersion σ_z, the adjustment cost parameter ϕ_ℓ, the demand parameters σ and ϵ, and the productivity disadvantage for new entrants d_E. To simplify the calibration procedure, I set the sunk cost of entry to 1. I

[10] An alternative way to generate a lifecycle of producer size is by assuming that entering producers face the same productivity distribution as incumbents but must pay adjustment frictions to grow to their initial optimal size. In Appendix E, I explore this alternative.

also assume $\mathbb{P}(\text{exit}) = 0.11$.

Table 1: Pre-set parameters

Parameter	Description	Value	Source/Target
β	Discount factor	0.96	Annual model
$\mathbb{P}(\text{exit})$	Probability of exit	0.11	Annual entry rate
ν	Inverse Frisch elasticity	0.5	Clementi and Palazzo (2016)
δ	Employment depreciation rate	0.19	Siemer (2014)

Note: This table summarizes part of the parameterization of the model. These parameter values were each chosen without targeting a particular moment in model simulations. Siemer (2014) estimates the employment depreciation rate as the average quit rate in JOLTS. Source: author's calculations.

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

Table 2: Calibrated parameters

Parameter	Description	Value	Targeted moment
ρ_s	TFP persistence	0.85	Frac. rel. sales below 1
σ_s	TFP innovation dispersion	0.15	Var. emp. growth
ϕ_ℓ	Adjustment cost	0.05	Autocorr. emp. growth
ϵ/σ	Superelasticity	0.67	Labor–sales regression
d_E	Prod. disadvantage of entrants	0.44	Average size entering producer
σ	Elasticity parameter	50	Average markup

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

The model performs well along a number of targeted and untargeted moments. Table 3 summarizes the model's fit. As in the data, the model generates a wedge between the variance of labor growth and the variance of sales growth. The wedge between these two numbers is in line with its value in the data. The model also fits the share of employment at entrant and young establishments that I estimate in the BDS. Fitting these variables is key to ensuring that the model accurately measures the aggregate importance of entrants. The model fits the distribution of producer size reasonably well, matching facts established in Edmond, Midrigan, and Xu (2018) using U.S. Census data. The model slightly overstates the fraction of producers with relative size below 1 and slightly understates the fraction with relative size below 10. The model misses the far right tail of producers with relative sales above 50, as shown in the last line of table 3.

Table 3: Calibration targets and model fit

Moment	Target	Source	Model moment
$\mathrm{Var}(\Delta \log L))$	0.06	Compustat	0.06
$\rho(\Delta \log L_t, \Delta \log L_{t-1})$	0.13	Compustat	0.14
Labor–sales regression	0.654	Compustat	0.665
Average size of entering producer	50 percent	CP	52 percent
Frac. rel. sales. below 1	87 percent	EMX	88 percent
Cost–weighted average markup	1.25	DLE	1.26
$\mathrm{Var}(\Delta \log PY))$	0.14	Compustat	0.14
$\rho(\Delta \log P_t Y_t, \Delta \log P_{t-1} Y_{t-1})$	0.12	Compustat	0.13
Frac. rel. sales below 10	99 percent	EMX	97 percent
Frac. rel. sales below 50	99.9 percent	EMX	100 percent

Note:DLE: De Loecker, Eeckhout, and Unger (2020), CP: Clementi and Palazzo (2016), EMX: Edmond, Midrigan, and Xu (2018)
Untargeted moments below line

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

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

Consistent with other studies that use microdata to estimate the superelasticity, the value of $\epsilon/\sigma = 0.67$ is nearly two orders of magnitude smaller than estimates using macroeconomic data. As noted by Klenow and Willis (2016), the large estimates of the superelasticity needed to account for macroeconomic persistence are inconsistent with micro–level evidence. In this model, setting the superelasticity near the estimates in Harding, Lindé, and Trabandt (2022) and Smets and Wouters (2007) would imply a counterfactually large markup-size relationship.

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

Robustness to superstars. While the baseline model captures some features of the heterogeneity in producer size, it does not match the extreme right tail of the firm size distribution in the U.S. data. In particular, it does not match the market share of the top 1% of producers. In the model it is 26%, whereas Compustat firms, a 1% sample of large firms in the US, account for 75% of sales in the US.

To study how this omission affects these results, in appendix F, I calibrate an alternative model with an additional "superstar" productivity state. I find that, in a calibration that matches the sales share of Compustat producers, the superelasticity falls by about half, to 0.33. This result suggests that the role of the variable elasticity of demand in amplifying entry shocks may be even smaller than I find in section 4.1.

3.3 Steady state producer dynamics

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

The regression coefficient in the model is 0.665. When I set $\phi_\ell = 0$, re-solve the model, simulate a panel of producers in the new model, and estimate the regression coefficient, I find $\hat{\beta}_L = 0.704$. When $\phi_\ell = 0.05$ (as in the baseline model) and the superelasticity of demand is 0, the regression coefficient rises to $\hat{\beta}_L = 0.941$. This decomposition suggests that labor adjustment costs account for between 12 percent and 20 percent of the deviation of the regression coefficient from 1.

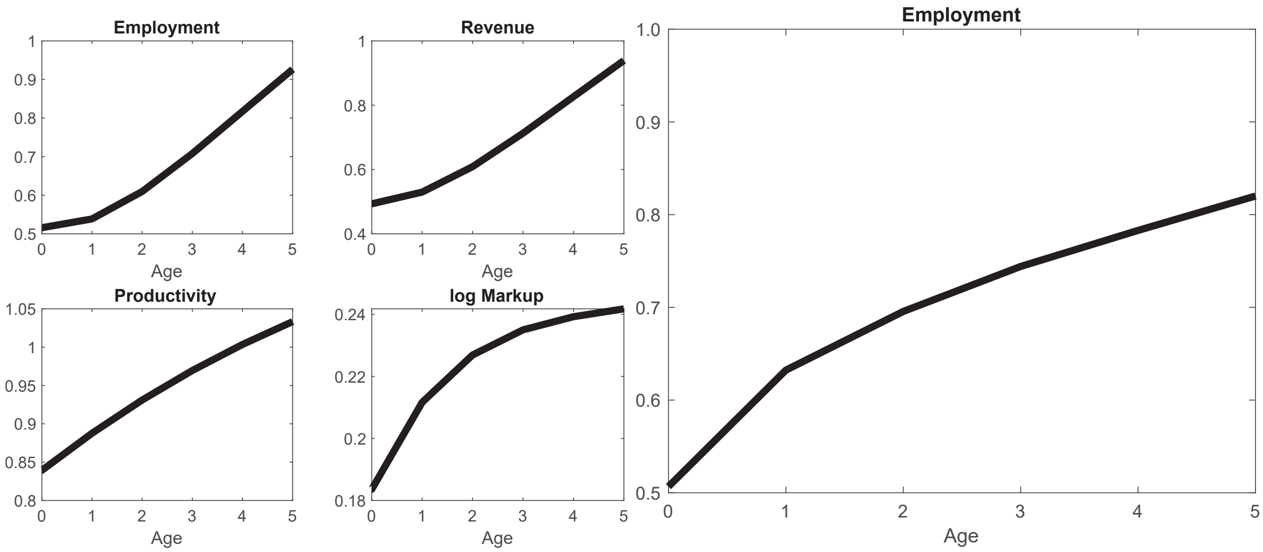
This decomposition shows that ignoring variable input adjustment costs would lead an econometrician to overstate the relationship between firm size and market power. However, it also shows that, even accounting for variable adjustment costs, large firms' markups do rise with with their market shares.

The lifecycle of the producer. Producers in the model, as in the data, begin small and grow slowly. Figure 3 shows that the average entering producer in the model employs around 50 percent of the labor force of the average incumbent producer, as in the data. In the model, they reach 90% of the size of the average producer by around age five. The model achieves this outcome in two ways: (1) the average productivity of entering producers is lower than that of incumbents and slowly reverts to the mean and (2) labor adjustment costs further slow the growth of new producers.

Producers' markups in the model also follow a life-cycle pattern, beginning low and slowly increasing.[11] The desire to set high markups derives from a demand elasticity that decreases with relative size. Because young producers grow slowly, their markups

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

Figure 3: Life cycle of the producer in the quantitative model and data



(A) Producer lifecycle in the model (B) Producer lifecycle in data

Note: The figure summarizes the lifecycle of an establishment in the model and in data. Each panel in subfigure A shows the path of the average of a particular establishment-level variable for producers of a particular age relative to its average for all incumbents. Subfigure B shows average employment relative to the population average in the BDS data. Source: U.S. Census Bureau, author's calculations.

also increase slowly with age. The cost–weighted average markup increases by around 6 log points over the first five years of a producer's life in the model.

Discussion. In this paper, I study a model with a collection of mechanisms, each of which is motivated by a feature of microdata. Tables 4 and 5 summarize these key mechanisms, whether they are present in other papers in the literature, and the features of the data that motivate those mechanisms. These mechanisms are: (1) variable input adjustment costs, (2) variable elasticity of demand, and (3) heterogeneity, including the producer lifecycle. This paper is the first study that combines all of these ingredients.

As Table 5 shows, a model missing any of these ingredients is at odds with the data. The adjustment cost in the model is key to generating the positive auto-correlation of net hiring present in microdata. Without any adjustment cost, the auto-correlation is negative, reflecting reversion to the mean in the productivity process. The variable elasticity of demand is key to generating the less than one-for-one relationship between sales and variable input growth that is present in the data. Without the variable elasticity, variable input use varies nearly one-for-one with sales. And lastly, the producer lifecycle is key to ensuring that entering producers represent a realistic share of aggregate employment.

The combination of these mechanisms differs from existing papers, as shown in Table 4. In the next section, I explore the role of each of these mechanisms for the propagation of entry fluctuations to the rest of the economy.

Table 4: Mechanisms present in quantitative theories of entry

Mechanism	This paper	Edmond, Midrigan, and Xu (2018)	Clementi and Palazzo (2016)	Bilbiie, Ghironi, and Melitz (2012)
Variable input adjustment cost	✓	✗	✗	✗
Variable markup	✓	✓	✗	✓
Producer lifecycle	✓	✗	✓	✗
Heterogeneous producers	✓	✓	✓	✗

Table 5: Key mechanisms and their identifying moments

Mechanism	Relevant Moment	Data value	Model value	Counterfactual value
Variable input adjustment cost	Auto-correlation of net hiring	0.13	0.14	-0.23
Variable markup	Regression of variable input bill growth on sales growth	0.654	0.665	0.941
Producer lifecycle	Avg. relative size entering producer	0.50	0.51	1.00

4 Market power, adjustment costs, heterogeneity, and entry fluctuations

The goal of this paper is to assess the role of entry fluctuations in amplifying business cycles. So far, I have described the quantitative structural framework I will use to make this assessment and presented a quantification of that framework. In this section, I study the response of the model economy to two different shocks: first, a shock to the cost of entry, and second, a shock to aggregate TFP. First studying the shock to the cost of entry isolates the effects of entry fluctuations from the other effects of the TFP shock, and it allows me to analyze the mechanisms that impact these effects. After discussing these mechanisms, I then analyze the role of entry in amplifying a TFP shock.

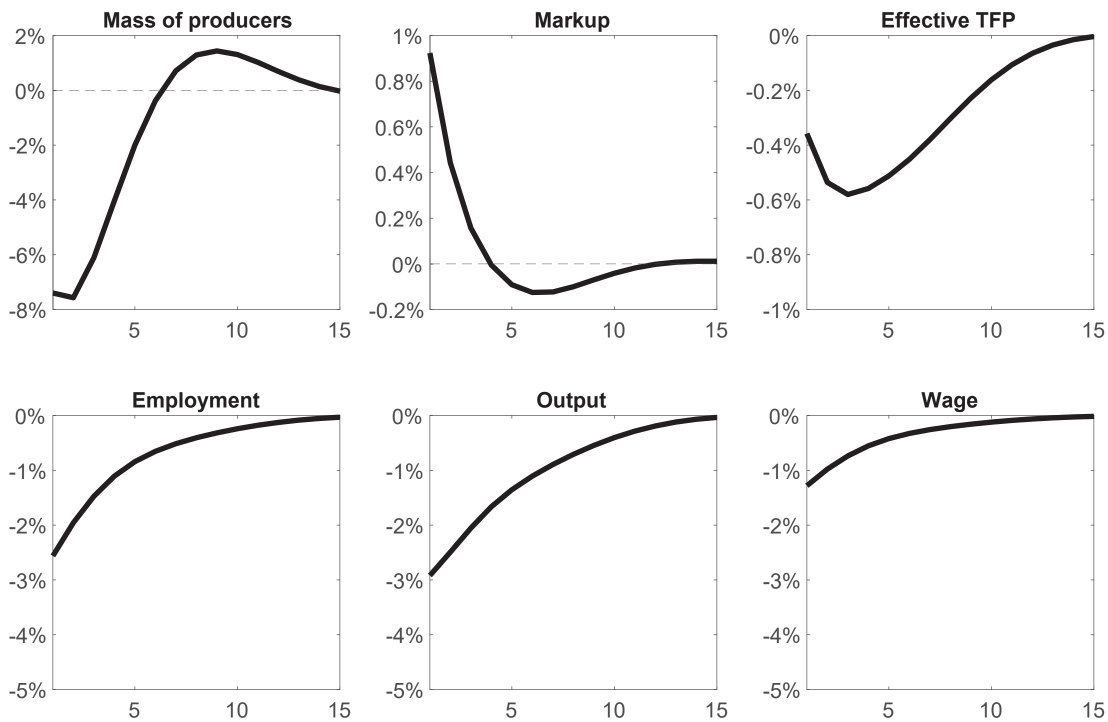
4.1 An entry shock

I first solve for the response of the model economy to an unexpected shock to the cost of entry. To isolate the effects of exogenous shocks to this cost, I impose that the entry cost varies over time but not endogenously; i.e., $c_E(\Lambda_t) = f_{E,t}$. I choose a path for $f_{E,t}$ so that the initial decline in the number of operating producers in the model matches the decline in the number of establishments during the Great Recession, and I assume that $f_{E,t}$ reverts back to its steady-state value with persistence 0.685.[12] After the initial shock is realized, all agents in the economy have perfect foresight of all aggregate variables as the economy returns to its steady state. I describe the solution method in detail in appendix B.

Figure 4 depicts the response of the model to a shock to the cost of entry. The shock causes a fall in entry that leads the mass of establishments to decline by a little over 7 percent. The cost–weighted average markup rises 0.9 percent. Effective TFP, equal to the ratio of output to aggregate employment, falls gradually by about 0.6 percent.

[12] This value is chosen to match the persistence of TFP shocks in Clementi and Palazzo (2016). The size of the initial increase in the cost of entry is roughly 8% of its steady state value.

Figure 4: Response of the baseline quantitative model to an entry cost shock



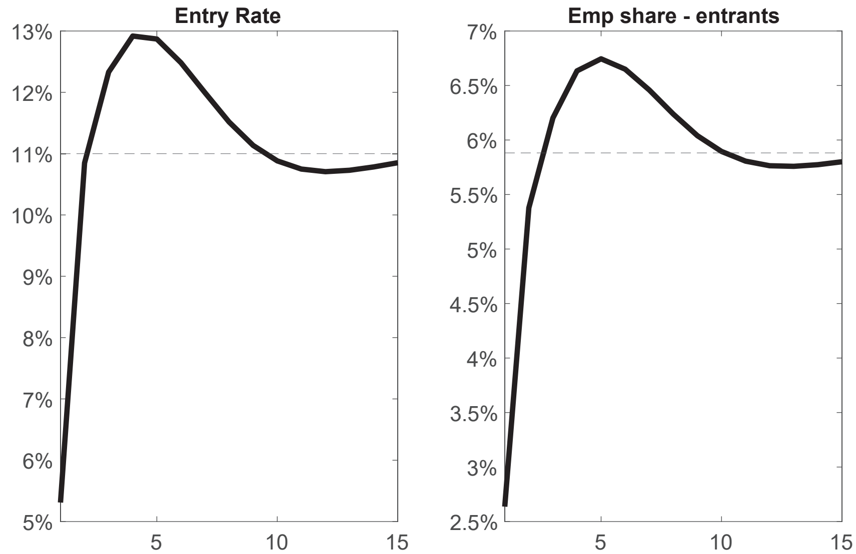
Note: The figure depicts the response of several aggregate variables to a persistent, unexpected shock to the cost of entry. Each line depicts the percent deviation of the variable from its steady state value. The size of the shock is chosen to match the fall in the number of establishments per capita during the Great Recession. Following the shock, the economy follows a perfect foresight path back to steady state. Source: author's calculations.

Employment falls about 2.5 percent on impact, and output falls nearly 3 percent. The wage satisfies the household labor supply equation and falls around 1 percent.

In response to the shock, the entry rate and share of employment among entrants fall. Figure 5 depicts the role of entrants following the shock. The entry rate falls by around 5 percentage points. It recovers quickly, with some overshooting, in part because the mass of entering producers (the numerator) recovers quickly while the mass of operating producers (the denominator) only gradually returns to its steady state level. The employment share among entering producers falls from 6 percent to around 3 percent. These fluctuations are all in line with those that the US experienced during the Great Recession.[13]

[13] According to the U.S. Census Bureau's BDS, the establishment entry rate fell from 13% to 9%, and the

Figure 5: Entry following the shock



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

4.2 Markups and TFP

To understand the roles of the average markup μ_t and effective TFP Z_t in generating the contraction in employment, it is useful to study the aggregated version of the model. This aggregated model is summarized by an aggregate production function (equation 23), the definition of the markup as the inverse labor share (equation 24), and the labor supply equation (equation 25).[14]

$$Y_t = Z_t L_t, \tag{23}$$

$$\mu_t = \frac{Y_t}{W_t L_t}, \tag{24}$$

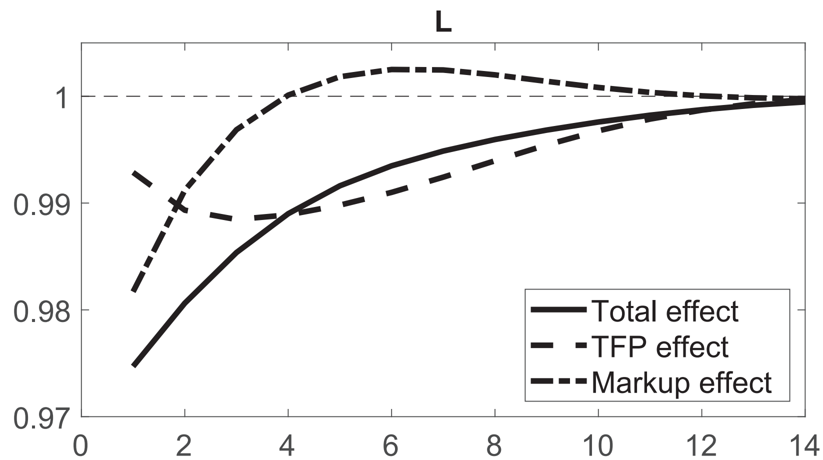
$$W_t = \psi L_t^{\nu}. \tag{25}$$

Given paths for the cost–weighted markup μ_t and aggregate effective productivity

share of employment among entering establishments fell from 5.5 percent to 3.5 percent.

[14] These can be further simplified into log-affine labor supply and labor demand equations: $\log W_t = \log \psi + \nu \log L_t$ and $\log W_t = \log Z_t - \log \mu_t$.

Figure 6: Decomposition of entry shock



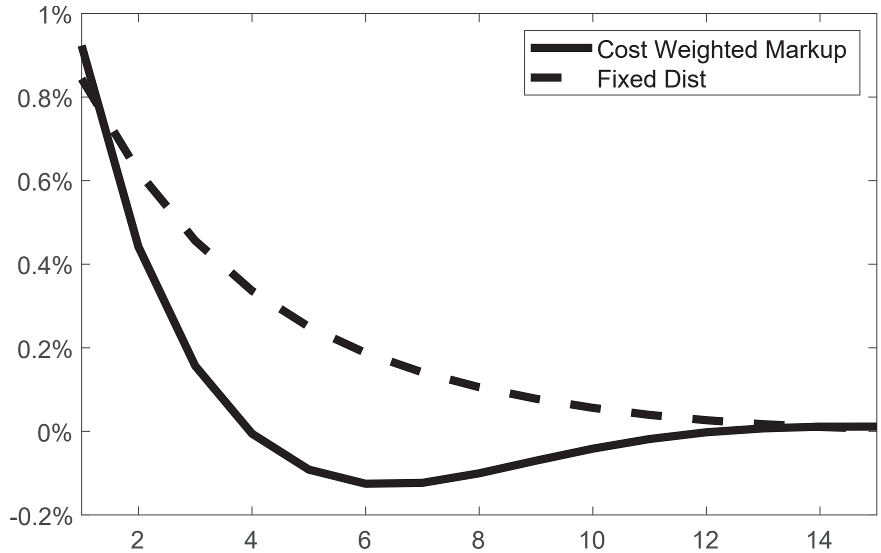
Note: The figure depicts a decomposition of the effects of the shock on aggregate employment into the effects of TFP and the effects of the markup. Each line depicts the contribution of either TFP or the markup to the percent deviation of employment from its steady state value. The dot-dashed line depicts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

Z_t, equations (23) to (25) imply paths for output Y_t, employment L_t, and the wage W_t. While changing the paths of μ_t or Z_t and recomputing these aggregate quantities does not necessarily constitute an equilibrium of this economy, this representation allows for a decomposition of the response of aggregate variables to a shock.

Figure 6 depicts the path of employment under different paths for the markup and productivity. In the solid line, I allow both to follow their equilibrium paths. In the dashed line, I hold the markup fixed, and, in the dot-dashed line, I hold TFP fixed. As they show, the rising markup generates a fall of 1.8 percent in employment, which represents about three-fourths of the immediate decline in employment. TFP accounts for the remainder of the decline in employment. Once the markup returns to its steady state value after a few years (with some overshooting), the decline in TFP accounts for all of the deviation in employment from steady state.

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

Figure 7: Reallocation and the markup



Note: This figure depicts the path of the cost-weighted average markup in response to the shock to the mass of potential entrants. The solid line depicts the path of the cost-weighted markup, allowing both the policy function of producers and the distribution of employment across producers to vary. The dashed line shows the markup, holding fixed the distribution of employment across producers. Source: author's calculations.

average markup:

$$\mathcal{M}_t = \int \int \mu_t(z,\ell) \frac{\ell_t(z,\ell)}{L_t} d\Lambda_t(z,\ell) \tag{26}$$

The decline in entry could affect the aggregate markup through changes in the markups of individual producers, $\mu_t(z)$, or the distribution of employment across producers, $\frac{\ell_t(z,\ell)}{L_t} d\Lambda_t(z,\ell)$. To understand the roles of each of these changes, figure 7 compares the path of the cost-weighted average markup (solid line) to the path of the cost-weighted average markup holding the distribution of employment fixed at its steady state (dashed line). As the path of the markup holding the distribution fixed shows, $\mu(z)$ increases on average; that is, the average producer raises its markups persistently in response to the shock. Allowing the distribution to vary, the cost-weighted average markup rises much less persistently than this counterfactual. This difference implies that the change in the distribution, $\frac{\ell_t(z,\ell)}{L_t} d\Lambda_t(z,\ell)$, reduces the markup following the shock, and on average, employment reallocates away from high markup producers and toward low markup producers.

There are two mechanisms in the model that lead producer-level markups to rise. First, the variable elasticity leads incumbent producers to raise their markups in response to the fall in entry; as entry falls, incumbents' market shares rise, and they increase their markups. Second, in the presence of adjustment costs, the shock pushes producers away from their static optima, which shows up as an increase in the markup. The decline in entry leads the wage to fall, increasing producers' labor demand, but the adjustment cost prevents them from immediately hiring to their optimal level. The wedge between the static optimum and actual employment shows up as a producer-level rise in the markup.

The shock leads employment to reallocate toward low-markup producers because of the variable elasticity of demand. In the model, small producers set lower markups because they face a higher elasticity of demand than large producers. Small producers also have higher pass-through. So, as the wage falls, small producers lower their prices by more than large producers, and since they face a higher elasticity of demand, they also grow relative to large producers. Thus, the decline in entry leads to a reallocation of employment to low-markup, high-elasticity producers.[15]

TFP and love for variety. Movements in effective TFP account for about a quarter of the initial decline in employment. There is love for variety in this model; effective TFP increases with the number of differentiated varieties. There is no closed-form expression for the love for variety effect in a model with heterogeneous firms and Kimball demand, so I instead compare the fall in aggregate TFP to the decline implied by the same reduction in the number of producers in a model of symmetric producers and CES demand. In that case, aggregate productivity is a function of the number of producers N and the elasticity of substitution σ_{CES}: $Z(N) = N^{\frac{1}{\sigma_{CES}-1}}$.

This calibration implies $\sigma_{CES} \approx 5$, so that love for variety implies a decline in effective TFP of around one quarter of the decline in the number of producers, or almost 2 percent.[16] This decline is much larger than the actual decline in effective

[15] A similar mechanism is present in Baqaee, Farhi, and Sangani (2021).

[16] I set σ_{CES} so that $\frac{\sigma_{CES}}{\sigma_{CES}-1}$ equals the cost-weighted markup in the benchmark model.

TFP. I return to this discrepancy in my discussion of the role of the producer life-cycle in Section 4.4.

4.3 The roles of adjustment costs and the variable elasticity of demand

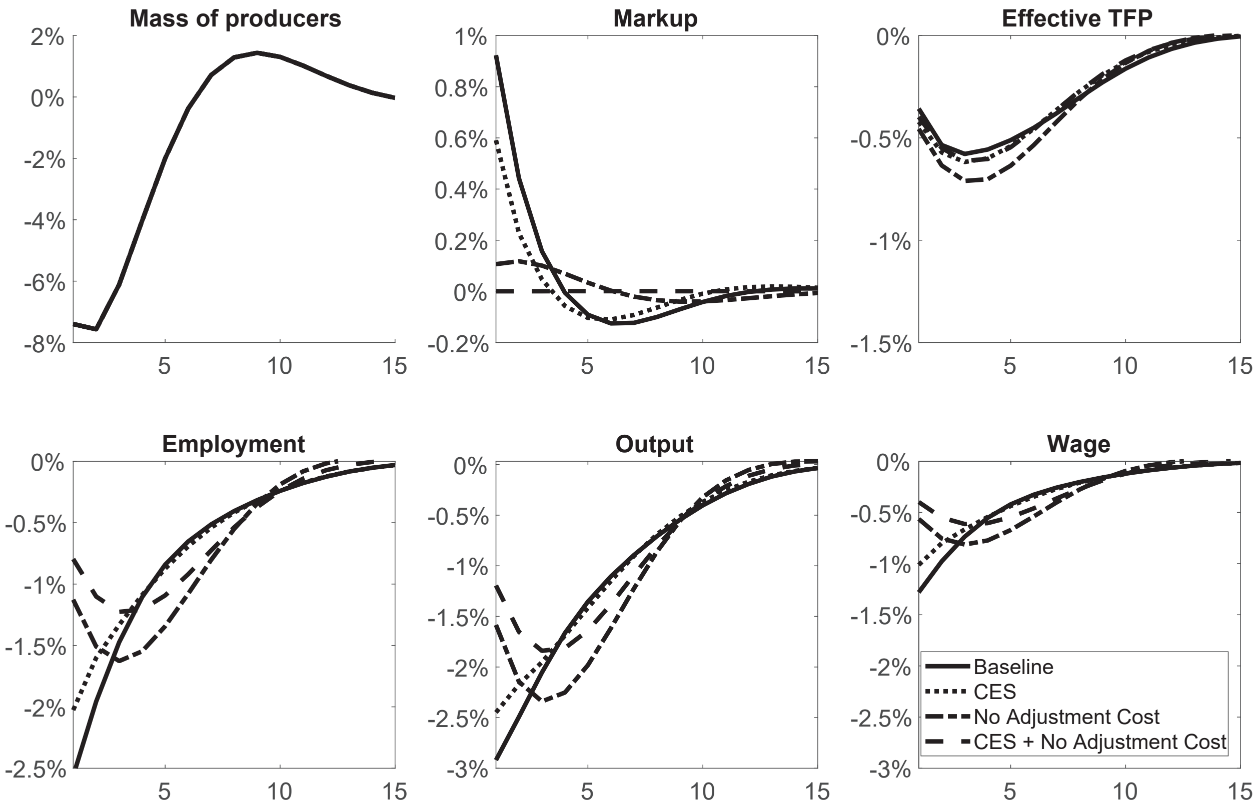
Two of the mechanisms identified in Table 5 are the variable elasticity of demand and the variable input adjustment cost. Figure 8 shows the response of the economy to the shock under different sets of assumptions concerning these two mechanisms. While both affect the propagation of the shock, I find that adjustment costs play a larger role: a model without the variable demand elasticity generates 80% of the decline in employment in the baseline model, while a model without adjustment costs generates only 50% of this decline.

The role of the variable elasticity of demand. To quantify the role of the variable elasticity, I compare the baseline model to one in which producers' demand elasticities do not vary. This alternative 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 steady-state cost–weighted markup in each model is identical. I keep all other parameters the same.

I subject each economy to the same shock as before. The line labeled "CES" in Figure 8 depicts the results of this experiment in the model with constant elasticity. As the figure shows, the CES model generates about 80 percent of the fall in employment and over 80 percent of the fall in output in the baseline model. The 20 percent difference in the employment and output responses between the two models arises because switching from a variable elasticity to a constant elasticity leads the aggregate markup to increase by about a third less. So, the variable elasticity of demand accounts for 20 percent of the overall decline in employment.

The role of adjustment costs. To quantify the role of adjustment frictions, I compare the baseline economy to one without adjustment costs. In the line denoted

Figure 8: Response to an entry cost shock in four economies



Note: This figure depicts the response of several aggregate variables to a persistent unexpected shock to the cost of entry in four different models. The solid line depicts the "baseline model" as described in Section 2. The dotted line depicts a model identical to the baseline except that it features a CES final goods production function, rather than Kimball. The dot-dashed line depicts a model that is identical to the baseline except that producers face no adjustment costs. Lastly, the dashed line depicts a model with CES production and no adjustment costs. Source: author's calculations.

"No Adjustment Cost" in figure 8, I show the response of this economy to the entry cost shock. As the figure shows, employment declines by about 50 percent less on impact in this model, driven by an increase in the markup that is almost 90 percent less than in the baseline model. So, adjustment costs account for about half of the employment response in the baseline economy.

Most of the difference between these two models can be accounted for by the aggregate markup, shown in the top middle panel of figure 8. As discussed above, adjustment costs have two effects on the markup in this model. First, in the presence of adjustment costs, the decline in the wage pushes producers away from their frictionless optimal solution. This wedge shows up in the model as a rise in the markup. And second, the variable elasticity of demand leads employment to reallocate to low-markup producers following the shock, and adjustment frictions slow this process.[17]

The model without adjustment costs is similar to the model in Edmond, Midrigan, and Xu (2018), who find that fluctuations in entry have little effect on the aggregate markup in a model with a variable elasticity, Pareto-distributed productivity, and frictionless adjustment of variable inputs. They find that an optimal entry subsidy, which increases the mass of operating firms by over 20%, has almost no effect on the aggregate markup. Consistent with their findings, this experiment shows that a shock to the cost of entry leads to a very small movement in the markup in a model with no adjustment frictions.

Model with constant markups and no adjustment costs. The lines labelled "CES + No Adjustment Cost" in figure 8 depict the response of the economy with CES demand and no adjustment costs to the entry cost shock. Without either adjustment costs or variable demand elasticity, the markup does not vary in response to the shock, leading employment to fall by 70 percent less than in the baseline economy.

[17] The importance of each of these channels can be read off of figure 8. Note that the initial rise in the markup in the Baseline economy relative to the "No Adjustment Cost" economy captures both channels. This difference is about 0.8 percentage point. On the other hand, the increase in the markup in the CES economy relative to the "CES + No Adjustment Cost" economy captures only the first channel. This difference is 0.6 percentage point. So, about three-fourths of the total markup response is due to the first channel, while the second channel accounts for the remaining one-fourth.

This model closely resembles that in Clementi and Palazzo (2016), who study the role of entry in amplifying aggregate TFP shocks in an industry equilibrium model with perfect competition, a firm lifecycle, and flexible labor adjustment. Consistent with their findings, this model implies that entry fluctuations have small but quite persistent effects on aggregate employment and output. However, the model with variable elasticity and adjustment costs also implies that a shock to entry has a larger effect on aggregate employment on impact.

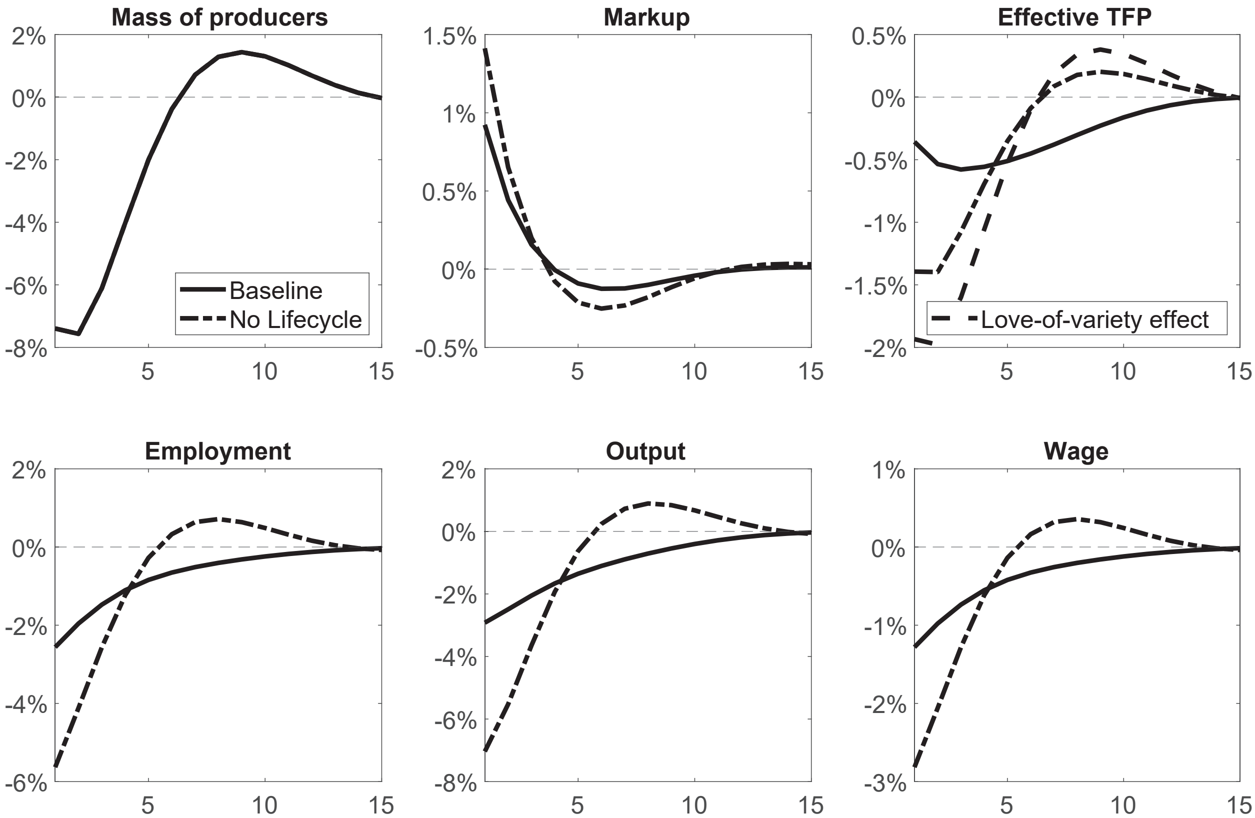
Discussion. In this section, I quantify the extent to which adjustment costs and the variable elasticity of demand amplify the employment effects of fluctuations in entry. I find that employment falls by 70 percent less in a model without either of these features than in the baseline model. Adjustment costs account for 50 percentage points of this difference, with the variable elasticity of demand accounting for the remaining 20 percentage points.

4.4 The role of heterogeneity

The firm lifecycle. Another mechanism identified in Table 5 is the lifecycle of the producer. As shown in Table 3, entering establishments are roughly half the size (in terms of employment) of the average establishment in the US. This fact is at odds with models that rely on the assumption of homogeneous producers, such as in Bilbiie, Ghironi, and Melitz (2012). To understand the role of the lifecycle in the propagation of entry fluctuations to the rest of the economy, I compare the baseline model to one in which entering producers are the same average size as incumbents.

The lines marked “No Lifecycle” in Figure 9 depict the results of this experiment. Employment initially falls by over twice as much in the economy with no lifecycle relative to the baseline, and it recovers much more quickly. The difference in the size of the impact and the speed of the recovery between these two impulses shows the role of the “missing cohort” effect; in the baseline model, because entering producers are small but grow over time, fluctuations in entry have a smaller but persistent effect on

Figure 9: The role of the lifecycle in the effects of an entry shock



Note: The figure depicts the response of several aggregate variables to a persistent, unexpected shock to the cost of entry. Each line in each panel depicts the percent deviation of the variable from its steady state value. The size of the shock is chosen to match the fall in the number of establishments per capita during the Great Recession. Following the shock, the economy follows a perfect foresight path back to steady state. Source: author's calculations.

aggregate employment. Inspecting the paths for the aggregate markup and effective TFP shows that most of the difference in the two employment responses is due to TFP. TFP falls by about 3 times as much in the economy with no lifecycle.

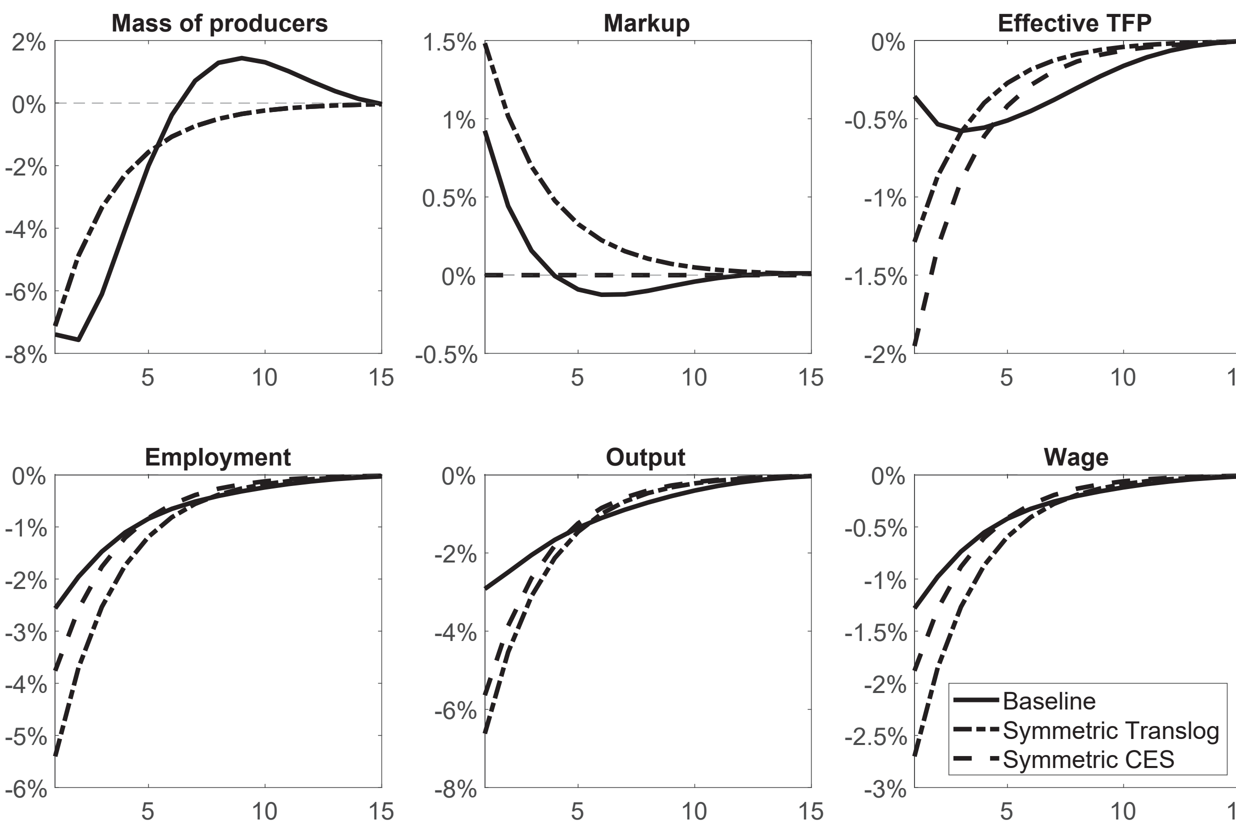
The dashed line labeled "love for variety effect" depicts the path of aggregate TFP under the symmetric CES formulation for the effects of love for variety. As it shows, the symmetric CES benchmark somewhat approximates the path of effective TFP in the model with no producer lifecycle. In the model with a lifecycle, love for variety effects are much smaller on impact because entrants are small and so contribute little to aggregate productivity.

Robustness. As Figure 3 shows, while the model matches the relative size of entrants, it overstates the speed at which new producers grow. In appendix H, I compare the baseline model to a calibration in which I use the adjustment cost to target the average employment of 5-year-old producers relative to incumbents. I find that the alternative calibration requires a significantly larger adjustment cost ($\phi_\ell = 0.15$) and so implies larger markup fluctuations than the baseline. It also implies smaller and more delayed effects on TFP, but on net, the markup effects dominate and this alternative model implies that entry matters more than in the baseline model.

A homogeneous producer model. Bilbiie, Ghironi, and Melitz (2012) study fluctuations in entry in a model with homogeneous producers and variable markups. To understand the relationship of my work to theirs, I compare the baseline model to two models in which all producers are identical, as in Bilbiie, Ghironi, and Melitz (2012). Following that paper, I consider models with CES and translog demand, and I set the labor adjustment cost $\phi_\ell = 0$. These models are otherwise identical to the baseline model I study. In the translog specification, the markup is a function of the number of producers: $\mu(N_t) = 1 + \frac{1}{\sigma N_t}$. Both models features love for variety. For more details, see appendix J.

I subject these symmetric economies to the same shock to the cost of entry that I study in the baseline exercises. The results of this experiment are depicted in Figure 10. In the symmetric translog model, employment falls by over twice as much as in the baseline model, reflecing an increase in the markup of 50% more and a decline in TFP that is three times as large. In the CES model, the TFP effects are even larger, though the markup does not move, resulting in employment effects that are between those in the baseline and symmetric translog model. These results show that accounting for heterogeneity greatly reduces the impact of a shock to the entry cost.

Figure 10: Entry and employment in symmetric models



Note: The figure depicts the response of several aggregate variables to a persistent, unexpected shock to the cost of entry. Each line in each panel depicts the percent deviation of the variable from its steady state value. The solid line shows this experiment in the baseline model. The dashed line shows the experiment in a symmetric CES model, and the dot-dashed line shows the experiment in a symmetric translog model. Source: author's calculations.

4.5 Entry and business cycle shocks

To study business cycles in the model, I modify the economy so that the cost of entry varies with the mass of operating producers; $\log c_E = \log f_E + \phi_E \log \frac{M_t}{\bar{M}}$, where M_t is the mass of operating producers at date t, and M is its steady state value. This modification allows me to use the parameter ϕ_E to target the extent to which the mass of producers varies with the aggregate TFP shock. If $\phi_E > 0$, then a decline in TFP that leads entry to decline also leads the entry cost to decline, dampening the fall in entry. I choose $\phi_E = 0.2$ so that, in the experiment below, the mass of operating producers declines as much as it did during the Great Recession.[18]

I subject the model economy to a shock to exogenous aggregate TFP, which I denote by A_t. The solid lines in figure 11 show the response of six model moments to a shock to TFP with a persistence of 0.685.[19] I choose the size of the shock so that aggregate employment falls by 8 percent, about how much the employment-population ratio fell during the Great Recession. In response to the exogenous decline in aggregate TFP, the mass of producers declines about 7 percent, around as much as the number of establishments per capita fell during the Great Recession. The markup rises 0.2 percent and effective TFP declines around 4 percent. Output falls around 12 percent and the wage falls roughly 4 percent.

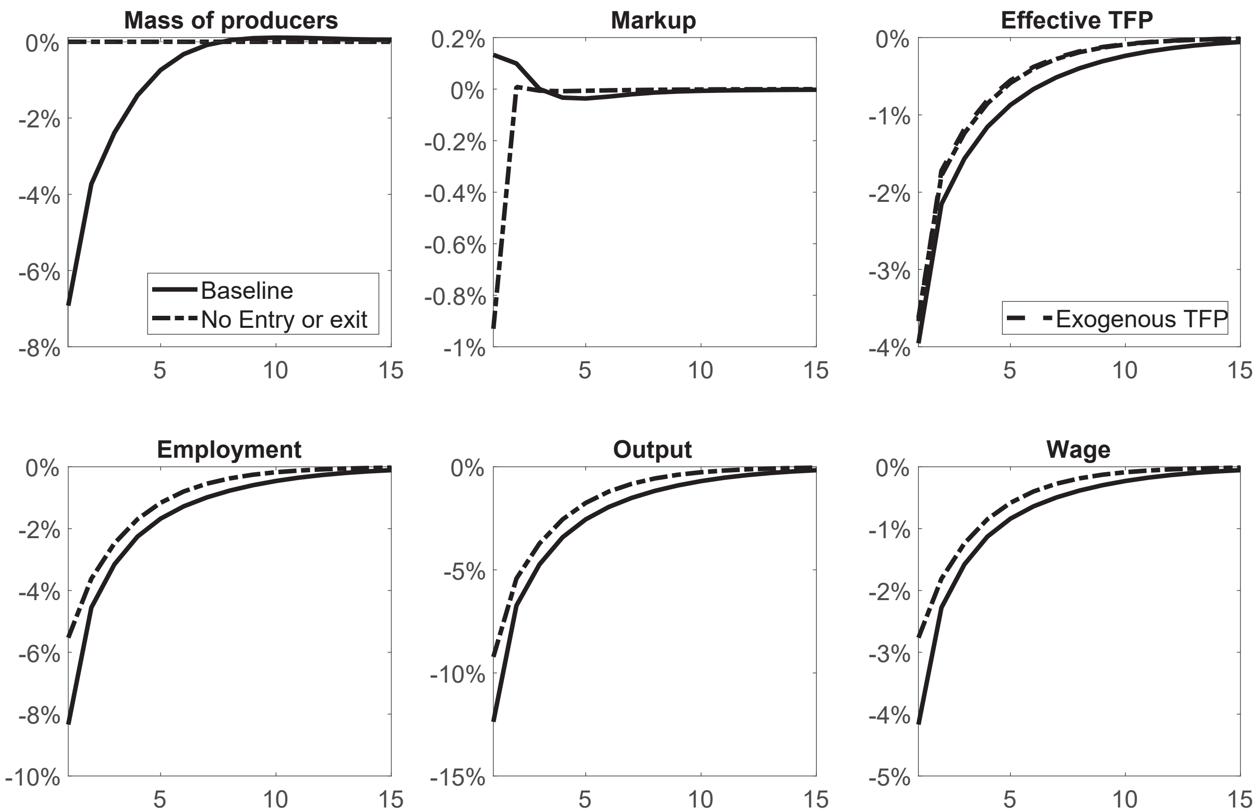
The role of entry. Figure 11 also shows the response of an economy without entry and exit to the same TFP shock.[20] As it shows, employment in the economy with entry and exit falls by around 50 percent more than in the economy without this margin of adjustment. Inspecting the other panels reveals that this difference is primarily due to the rise in the markup in the baseline model, which contrasts with a fall in the markup in the model without entry and exit. The rise in the markup in the baseline economy

[18] A similar mechanism is present in Gutiérrez, Jones, and Philippon (2021). In Appendix I, I explore alternative specifications for the entry cost. I find that specifications in which the entry cost varies less with the business cycle imply a larger role for entry. For example, I find that if $\phi_E = 0$, the mass of producers declines by more than it did during the Great Recession, likely overstating the role of endogenous fluctuations in entry during that period.

[19] I choose the persistence of the shock to follow Clementi and Palazzo (2016).

[20] To keep the two models comparable, I set the mass of producers in the "no entry" economy equal to the steady state mass of producers in the baseline model.

Figure 11: The role of entry in TFP shocks



Note: The figure depicts the path of the cost-weighted average markup in response to the shock to aggregate TFP in the baseline model and in one without entry and exit. Each panel depicts the percent deviation of the variable to its steady state value. Source: author's calculations.

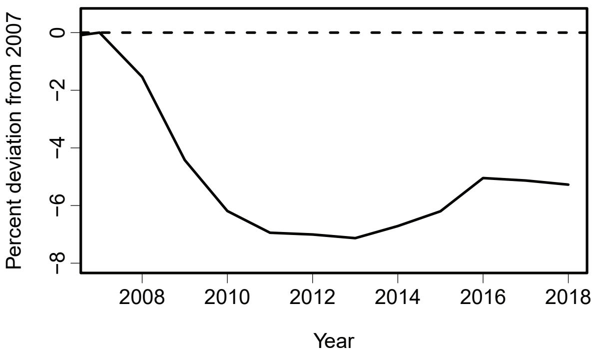
relative to the decline in the economy without entry and exit is roughly in line with the rise in the markup in response to the shock to the cost of entry discussed in Section 4.1.[21]

Great Recession. I chose the size of the TFP shock to match the decline in employment relative to trend during the Great Recession, and it generates a decline in entry consistent with the data during the same period. Comparing the path of employment in the baseline and "no entry or exit" models suggests that fluctuations in entry account for 2.5 percentage points of the 8 percent decline in employment during that period.

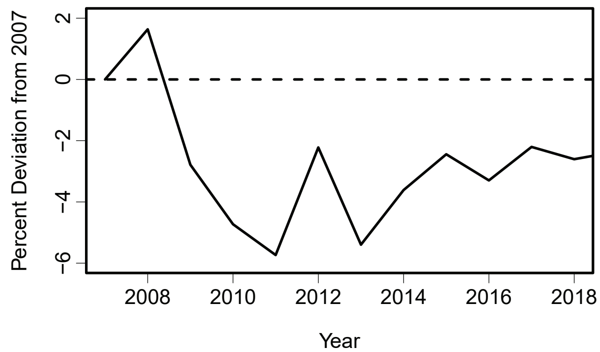
Figure 12 depicts the path of three model outcomes in the data during the Great Recession. As it shows, the employment-population ratio fell by about 8 percent during the Great Recession, with a slow recovery thereafter. By construction, employment in the model falls by just as much. The number of establishments per capita declines by a bit over 7 percent, about as much as it declines in the model. Lastly, the labor share declines by much more in the data than in the model. In the model, the markup rises by 0.2 percent, which implies a 0.2 percent fall in the labor share, whereas the labor share declines by an order of magnitude larger in the data, declining by almost 6 percent between 2007 and 2011.

[21] The markup in the model without entry and exit falls because of adjustment costs. To see why, recall that the marginal cost of a producer in this model equals $\frac{W_t}{z_t A_t}$, where A_t is exogenous, aggregate TFP. In the "no entry" model, both W_t and A_t decline, but A_t declines by more than W_t, leading marginal costs to rise. Producers do not fully and immediately change their labor demand following this shock, so markups fall temporarily. Note that while the markup declines in this model, it rises following the shock to entry. The reason for these opposite effects is that the TFP shock leads marginal costs to rise, whereas the shock to entry leads only the wage to decline, so marginal costs fall, leading the markup to rise.

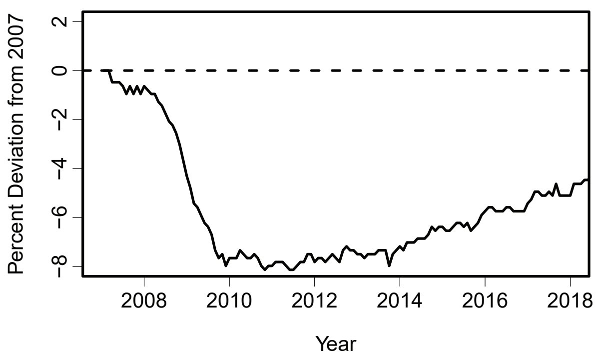
Figure 12: The Great Recession in Data



(A) Establishments per capita



(B) Labor Share



(C) Employment-population ratio

Note: Each panel depicts the path of a variable relative to its value in 2007. Panel A: Establishments per capita. Source: Establishments data come from the US Census Bureau Business Dynamics Statistics Database. Population data: U.S. Bureau of Economic Analysis, retrieved from FRED, Federal Reserve Bank of St. Louis; Panel B: Data on the US Labor Share come from the U.S. Bureau of Labor Statistics, retrieved from FRED; Panel C: Employment-population ratio data come from the U.S. Bureau of Labor Statistics retrieved from FRED.

5 Conclusion

In this paper, I assess the role of entry fluctuations in amplifying recessions in a general equilibrium model. The model features heterogeneous producers who face adjustment frictions and a variable elasticity of demand. My main finding is that entry plays an economically meaningful role in business cycle amplification, in large part due to the role of adjustment frictions. I also show that heterogeneity meaningfully reduces these effects.

This paper follows a substantial literature studying the role of entry fluctuations in business cycle amplification. Papers in that literature come to very different conclusions, depending on the assumptions underlying those results. In this paper, I provide a framework for understanding those disparate results. In particular, I show that models that ignore heterogeneity likely overstate the role that entry plays in business cycles, and models with heterogeneity that omit either a variable elasticity of demand or adjustment frictions likely understate the role that it plays.

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

Data Availability

Data and code replicating the tables and figures of this paper can be found in Gamber (2023) in the Harvard DataVerse, https://doi.org/10.7910/DVN/AWYAP2. Parts of the analysis in this paper use proprietary Compustat data; instructions for obtaining these data are contained in the replication package.

References

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

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

Baqaee, David, Emmanuel Farhi, and Kunal Sangani. 2021. *The Supply-Side Effects of Monetary Policy.* Working Paper, Working Paper Series 28345. National Bureau of Economic Research.

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

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

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics* 121:1–14.

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

Davis, Steven J, John Haltiwanger, and Scott Schuh. 1996. "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts." *Small Business Economics* 8 (4): 297–315.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *The Quarterly Journal of Economics* 135, no. 2 (January): 561–644.

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

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

Gamber, William. 2023. "Replication Data for: The Importance of Entry in the Business Cycle: What are the Roles of Markups, Adjustment Costs, and Heterogeneity?" *Harvard Dataverse,* https://doi.org/10.7910/DVN/AWYAP2.

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

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

Gutiérrez, Germàn, Callum Jones, and Thomas Philippon. 2021. "Entry costs and aggregate dynamics." *Journal of Monetary Economics* 124:S77–S91.

Harding, Martín, Jesper Lindé, and Mathias Trabandt. 2022. "Resolving the missing deflation puzzle." *Journal of Monetary Economics* 126:15–34.

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

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

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

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

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

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

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

Winberry, Thomas. 2021. "Lumpy Investment, Business Cycles, and Stimulus Policy." *American Economic Review* 111 (1): 364–96.