

# Entry, Exit, Firm Dynamics, and Aggregate Fluctuations\*

Gian Luca Clementi<sup>†</sup>      Berardino Palazzo<sup>‡</sup>

NYU and NBER      Boston University

Forthcoming on the American Economic Journal: Macroeconomics

## Abstract

Do firm entry and exit play a major role in shaping aggregate dynamics? Our answer is yes. Entry and exit amplify and propagate the effects of aggregate shocks. In turn, this leads to greater persistence and unconditional variation of aggregate time series. These results stem from well-documented features of firm dynamics such as pro-cyclical entry and the negative association between age and growth. In the aftermath of a positive aggregate productivity shock, the number of entrants increases. Since the new firms are smaller than the incumbents, as in the data, the initial impact on aggregate dynamics is negligible. However, as the common productivity component reverts to its unconditional mean, the new entrants that survive grow larger over time, generating a wider and longer expansion than in a scenario without entry or exit. The theory also identifies a causal link between the drop in establishments at the outset of the great recession and the painstakingly slow speed of the recovery from it.

**Key words.** Great Recession, Selection, Propagation, Persistence, Survival, Reallocation.

JEL Codes: E13, E22, E23, E32 D21, D92.

---

\*We are especially grateful to Richard Rogerson for suggestions that greatly improved the paper. We also acknowledge insightful comments from Fernando Alvarez, Dave Backus, Nick Bloom, Jeff Campbell, Russell Cooper, Jason Faberman, Shigeru Fujita, Fabio Ghironi, François Gourio, Ramon Marimon, Ben Pugsley, Gianluca Violante, and Stan Zin, as well seminar attendants at Banque de France, Boston College, Boston University, CUNY, European University Institute, North Carolina State, New York University, Northwestern University, Richmond Fed, University of Connecticut, University of Southern California, University of Texas at Austin, University of Virginia, University of Western Ontario, NBER EFG Meeting, Stockman Conference, AEA meeting, SED meeting, Canadian Macroeconomics Study Group, CREI-MOVE Workshop on Misallocation and Productivity, Greenline Macroeconomics Meeting, NY/Philly Macroeconomics Workshop, and Workshop on Macroeconomic Dynamics for their comments and suggestions. All remaining errors are our own responsibility.

<sup>†</sup>Department of Economics, Stern School of Business, New York University, [NBER](#), and [RCEA](#). Email: [clem@nyu.edu](mailto:clem@nyu.edu). Web: <http://people.stern.nyu.edu/gclement/>

<sup>‡</sup>Department of Finance, Questrom School of Business, Boston University. Email: [bpalazzo@bu.edu](mailto:bpalazzo@bu.edu). Web: <http://people.bu.edu/bpalazzo/>

# 1 Introduction

Over the last 25 years, empiricists have pointed out a tremendous amount of between-firm and between-plant heterogeneity, even within narrowly defined sectors. A key issue in the macroeconomics literature is to gauge the importance of such heterogeneity for the evolution of aggregate magnitudes.<sup>1</sup> The objective of this paper is to assess the role that entry and exit dynamics play in the propagation of aggregate shocks.

Our framework builds upon [Hopenhayn \(1992\)](#)'s model, amended to allow for investment in physical capital and for aggregate fluctuations. We posit that firms' productivity is the product of a common and of an idiosyncratic component, both driven by persistent stochastic processes and whose innovations are orthogonal to each other. The demand for firms' output and the supply of physical capital are infinitely elastic at the unit price, while the supply of labor services has finite elasticity. The wage rate fluctuates to ensure that the labor market clears.

When parameterized to match a set of empirical regularities on investment, entry, and exit, our framework replicates well-documented stylized facts about firm dynamics. The exit hazard rate declines with age. Employment growth decreases with size and age, both unconditionally and conditionally. The size distribution of firms is skewed to the right and the skewness of a cohort's size distribution declines with age. Furthermore, the entry rate is pro-cyclical, while the exit rate is counter-cyclical.

Our analysis revolves around two exercises. First, we consider the response of our economy to a positive aggregate TFP shock. Then, motivated by the large drop in entry that characterized the last recession, we go on to assess the role that a negative shock specific to the entry margin may have played in shaping aggregate dynamics since 2008.

We find that the effect of the entry and exit margins on aggregate dynamics is quantitatively relevant. Hitting our economy with a shock to aggregate productivity equal to 1.5 standard deviations leads to a 2.97% increase in cumulative output – with respect to trend – over the next 10 years. Performing the same experiment on a version of our model without entry or exit only elicits a 2.46% average cumulative output growth. In other words, entry and exit are responsible for almost one fifth of output growth over the 10 years following the shock.

Entry and exit also boost persistence. For the version without entry or exit to generate a data-conforming persistence of output, the first-order autocorrelation of aggregate productivity shocks must be as high as 0.775. In the benchmark scenario with entry and exit,

---

<sup>1</sup>For example, see the work of [Veracierto \(2002\)](#), [Khan and Thomas \(2008\)](#), [Bachman and Bayer \(2013\)](#), and [Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry \(2014\)](#).

it only needs to be 0.685. As pointed out by Cogley and Nason (1995), many competitive macroeconomic models have weak internal propagation mechanisms. In order to generate the persistence in aggregate time series that we recover in the data, they must rely heavily on external sources of dynamics. Our work shows that allowing for firm heterogeneity and for entry and exit can sensibly reduce such reliance.

What accounts for the propagation effect of entry and exit? A positive shock to the common productivity component makes entry more appealing. This is the case because, with a labor supply elasticity calibrated to match the standard deviation of employment relative to output, the equilibrium response of the wage rate is not large enough to undo the impact of the innovation to aggregate productivity.

Entrants are more plentiful, but of lower average idiosyncratic efficiency. This happens because, with a time-invariant cost of entry, firms with poorer prospects about their productivity find it worth entering. Everything else equal, this selection effect reduces aggregate output and TFP. However, given the small output share of entering firms, the contemporaneous effect is small.

It is the evolution of the new entrants that causes a sizeable impact on aggregate dynamics. As the common productivity component declines towards its unconditional mean, there is a larger-than-average pool of young firms that increases in size and efficiency. The distribution of firms over idiosyncratic productivity improves. It follows that entry propagates the effects of aggregate productivity shocks on output and increases its unconditional variance.

While we chose to illustrate its equilibrium properties by means of a positive innovation to productivity, our model has symmetrical implications for the role played by entry and exit during recessions. According to our theory, recoveries from troughs of economic activity are dragged down by a smaller-than-average fraction of young, high-growth firms. Given the unusually low entry rates recorded in 2008 and 2009, it is natural to ask whether such demographics contributed to slowing down the recovery from the great recession.

For our model to generate a decline in operating establishments of the magnitude recorded at the outset of the recession, an empirically sensible negative aggregate shock must be supplemented by a temporary shock to entrants' productivity. We find that the entry-specific shock has a very limited contemporaneous effect on output and employment, but leads indeed to a substantially slower recovery. Over time, the entry-specific shock becomes relatively more important, since the number of operating establishment mean-reverts at a slower pace than the exogenous component of aggregate TFP.

We are not the first to assess the role of entry and exit in a model with aggregate fluctuations. In particular, we are not the first to find that entry and exit enhance the effects of aggregate shocks. Devereux, Head, and Lapham (1996), Chatterjee and Cooper (2014), Bilbiie, Ghironi, and Melitz (2012), and Jaimovic and Floetotto (2008) model entry and exit in general equilibrium models with monopolistic competition. In the former three, a surge in entry leads to greater diversity of the product space. Because of increasing returns, this encourages agents to work harder and accumulate more capital. In Jaimovic and Floetotto (2008), more entry means more competition and lower markups.

Our approach is different, in that our results do not rely on either increasing returns, counter-cyclical market power, or other inefficiencies. In fact, our economy is efficient. The amplification and propagation results depend on the pro-cyclicality of the entry rate, for which evidence abounds,<sup>2</sup> and on the dynamics of young firms. Consistent with the evidence – see Foster, Haltiwanger, and Krizan (2001) – in our theory the relative importance of a cohort is minimal at birth and increases over time.<sup>3</sup>

The four papers mentioned above abstract completely from firm dynamics. Our research shows that this methodological choice is not without loss of generality. Carefully modeling the firm life-cycle sheds light on an important channel through which firm demographics affect aggregate dynamics. Recent work by Choi (2014) and Sedlacek (2015) also stress the role that the evolution in the firm distribution may play in shaping the recovery from economic recessions.

The remainder of the paper is organized as follows. The model is introduced in Section 2. In Section 3, we characterize firm dynamics in the stationary economy. The analysis of the scenario with aggregate fluctuations begins in Section 4, where we describe the impact of aggregate shocks on the entry and exit margins. In Section 5, we characterize the cyclical properties of entry and exit rates, as well as the relative size of entrants and exiters. We also gain insights into the mechanics of the model by describing the impulse responses to an aggregate productivity shock and argue that our results generalize to environments with alternative but empirically sensible assumptions on the cyclicality of the stochastic discount factor and on the exact nature of idiosyncratic shocks. In Section 6, we illustrate how allowing for entry and exit strengthens the model’s internal propagation mechanism and we outline its implications for the cyclicality of the cross-sectoral distribution of productivity. In Section 7, we show that when the model is consistent with a key feature

---

<sup>2</sup>See Lee and Mukoyama (2015).

<sup>3</sup>Three papers – Campbell (1998), Samaniego (2008), and Lee and Mukoyama (2013) – share our choice of sticking to competitive, efficient economies. None of them, however, identify or study the mechanism at the core of our contribution.

of the last recession, i.e. the large drop in the establishment count, it can also account for another noticeable property – the slow recovery. Section 8 concludes.

## 2 Model

Time is discrete and is indexed by  $t = 1, 2, \dots$ . The horizon is infinite. At time  $t$ , a positive mass of price-taking firms produce an homogenous good by means of the production function  $y_t = z_t s_t (k_t^\alpha l_t^{1-\alpha})^\theta$ , with  $\alpha, \theta \in (0, 1)$ .<sup>4</sup> With  $k_t$  we denote physical capital,  $l_t$  is labor, and  $z_t$  and  $s_t$  are aggregate and idiosyncratic random disturbances, respectively.

The common component of productivity  $z_t$  is driven by the stochastic process

$$\log z_{t+1} = \rho_z \log z_t + \sigma_z \varepsilon_{z,t+1},$$

where  $\varepsilon_{z,t} \sim N(0, 1)$  for all  $t \geq 0$ . The dynamics of the idiosyncratic component  $s_t$  is described by

$$\log s_{t+1} = \rho_s \log s_t + \sigma_s \varepsilon_{s,t+1},$$

with  $\varepsilon_{s,t} \sim N(0, 1)$  for all  $t \geq 0$ . The conditional distribution of  $s_t$  will be denoted as  $H(s_{t+1}|s_t)$ .

Firms hire labor services on the spot market at the wage rate  $w_t$  and discount future profits by means of the time-invariant factor  $\frac{1}{R}$ ,  $R > 1$ . Adjusting the capital stock by  $x$  bears a cost  $g(x, k)$ . Capital depreciates at the rate  $\delta \in (0, 1)$ .

We assume that the demand for the firm's output and the supply of capital are infinitely elastic and normalize their prices at 1. The supply of labor is given by the function  $L_s(w) = w^\gamma$ , with  $\gamma > 0$ .

Operating firms incur a cost  $c_f > 0$ , drawn from the common time-invariant distribution  $G$ . Firms that quit producing cannot re-enter the market at a later stage and recoup the undepreciated portion of their capital stock, net of the adjustment cost of driving it to 0. The timing is summarized in Figure 1.

Every period there is a constant mass  $M > 0$  of prospective entrants, each of which receives a signal  $q$  about her productivity, with  $q \sim Q(q)$ . Conditional on entry, the distribution of the idiosyncratic shock in the first period of operation is  $H(s'|q)$ , strictly decreasing in  $q$ . Entrepreneurs that decide to enter the industry pay an entry cost  $c_e \geq 0$ .

At all  $t \geq 0$ , the distribution of operating firms over the two dimensions of heterogeneity is denoted by  $\Gamma_t(k, s)$ . Finally, let  $\lambda_t \in \Lambda$  denote the vector of aggregate state variables and  $J(\lambda_{t+1}|\lambda_t)$  its transition operator. In Section 4, we will show that  $\lambda_t = \{\Gamma_t, z_t\}$ .

---

<sup>4</sup>While the calibration to follow relies exclusively on plant-level data, throughout the paper we will alternatively refer to our production units as firms or plants.

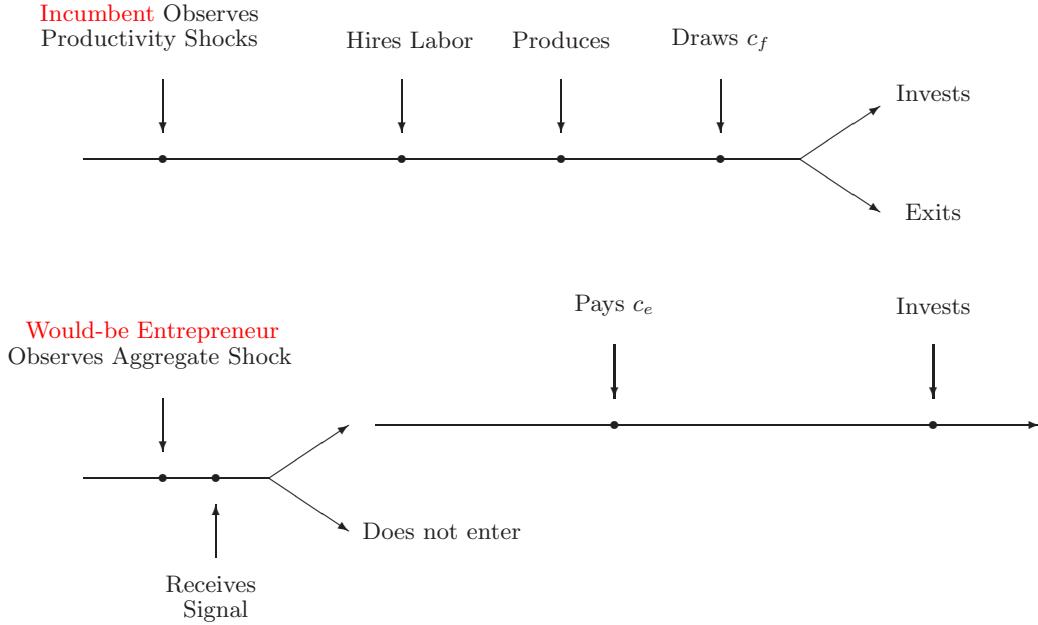


Figure 1: Timing in period  $t$ .

## 2.1 The incumbent's optimization program

Given the aggregate state  $\lambda$ , capital in place  $k$ , and idiosyncratic productivity  $s$ , the employment choice is the solution to the following static problem:<sup>5</sup>

$$\pi(\lambda, k, s) = \max_l sz[k^\alpha l^{1-\alpha}]^\theta - wl.$$

Upon exit, a firm obtains a value equal to the undepreciated portion of its capital  $k$ , net of the adjustment cost it incurs in order to dismantle it, i.e.  $V_x(k) = k(1-\delta) - g[-k(1-\delta), k]$ .

Then, the start-of-period value of an incumbent firm is dictated by the function  $V(\lambda, k, s)$  which solves the following functional equation:

$$V(\lambda, k, s) = \pi(\lambda, k, s) + \int \max \left\{ V_x(k), \tilde{V}(\lambda, k, s) - c_f \right\} dG(c_f)$$

$$\begin{aligned} \tilde{V}(\lambda, k, s) &= \max_x -x - g(x, k) + \frac{1}{R} \int_{\Lambda} \int_{\mathfrak{R}} V(\lambda', k', s') dH(s'|s) dJ(\lambda'|\lambda) \\ \text{s.t. } k' &= k(1-\delta) + x. \end{aligned}$$

## 2.2 Entry

For an aggregate state  $\lambda$ , the value of a prospective entrant that obtains a signal  $q$  is

$$V_e(\lambda, q) = \max_{k'} -k' + \frac{1}{R} \int_{\Lambda} \int_{\mathfrak{R}} V(\lambda', k', s') dH(s'|q) dJ(\lambda'|\lambda).$$

She will invest and start operating if and only if  $V_e(\lambda, q) \geq c_e$ .

---

<sup>5</sup>We drop time indexes in the remainder of the paper, except in parts where such choice may jeopardize clarity of exposition.

### 2.3 Recursive Competitive Equilibrium

For given  $\Gamma_0$ , a recursive competitive equilibrium consists of (i) value functions  $V(\lambda, k, s)$ ,  $\tilde{V}(\lambda, k, s)$  and  $V_e(\lambda, q)$ , (ii) policy functions  $x(\lambda, k, s)$ ,  $l(\lambda, k, s)$ ,  $k'(\lambda, q)$ , and (iii) bounded sequences of wages  $\{w_t\}_{t=0}^\infty$ , incumbents' measures  $\{\Gamma_t\}_{t=1}^\infty$ , and entrants' measures  $\{\mathcal{E}_t\}_{t=0}^\infty$  such that, for all  $t \geq 0$ ,

1.  $V(\lambda, k, s)$ ,  $\tilde{V}(\lambda, k, s)$ ,  $x(\lambda, k, s)$ , and  $l(\lambda, k, s)$  solve the incumbent's problem;
2.  $V_e(\lambda, q)$  and  $k'(\lambda, q)$  solve the entrant's problem;
3. The labor market clears:  $\int l(\lambda_t, k, s) d\Gamma_t(k, s) = L_s(w_t) \quad \forall t \geq 0$ ,
4. For all Borel sets  $\mathcal{S} \times \mathcal{K} \subset \Re^+ \times \Re^+$  and  $\forall t \geq 0$ ,

$$\mathcal{E}_{t+1}(\mathcal{S} \times \mathcal{K}) = M \int_{\mathcal{S}} \int_{\mathcal{B}_e(\mathcal{K}, \lambda_t)} dQ(q) dH(s'|q),$$

where  $\mathcal{B}_e(\mathcal{K}, \lambda_t) = \{q \text{ s.t. } k'(\lambda_t, q) \in \mathcal{K} \text{ and } V_e(\lambda_t, q) \geq c_e\}$ ;

5. For all Borel sets  $\mathcal{S} \times \mathcal{K} \subset \Re^+ \times \Re^+$  and  $\forall t \geq 0$ ,

$$\Gamma_{t+1}(\mathcal{S} \times \mathcal{K}) = \int_{\mathcal{S}} \int \int_{\mathcal{B}(\mathcal{K}, \lambda_t, c_f)} d\Gamma_t(k, s) dG(c_f) dH(s'|s) + \mathcal{E}_{t+1}(\mathcal{S} \times \mathcal{K}),$$

where  $\mathcal{B}(\mathcal{K}, \lambda_t, c_f) = \{(k, s) \text{ s.t. } \tilde{V}(\lambda_t, k, s) - c_f \geq V_x(k) \text{ and } k(1 - \delta) + x(\lambda_t, k, s) \in \mathcal{K}\}$ .

## 3 The Stationary Case

We begin by analyzing the case without aggregate shocks, i.e.  $\sigma_z = 0$ . In this scenario, our economy converges to a stationary equilibrium in which all aggregate variables are constant. First, we introduce our assumptions about functional forms. Then we illustrate the mechanics of entry, investment, and exit. After calibrating the model, we detail the model's implications for several features of firm dynamics, among which the relation between firm growth and survival, size, and age.

### 3.1 Functional Forms

Investment adjustment costs are the sum of a fixed portion and of a convex portion:

$$g(x, k) = \chi(x)c_0k + c_1 \left(\frac{x}{k}\right)^2 k, \quad c_0, c_1 \geq 0,$$

where  $\chi(x) = 0$  for  $x = 0$  and  $\chi(x) = 1$  otherwise. Notice that the fixed portion is scaled by the level of capital in place and is paid if and only if gross investment is different from zero.

The distribution of signals for the entrants is Pareto. See the left panel of Figure 2. We posit that  $q \geq \underline{q} \geq 0$  and that  $Q(q) = (\underline{q}/q)^\xi$ ,  $\xi > 1$ . The realization of the idiosyncratic shock in the first period of operation follows the process  $\log(s) = \rho_s \log(q) + \sigma_s \eta$ , where  $\eta \sim N(0, 1)$ .

Finally, we assume that the distribution of the operating cost  $G$  is log-normal with parameters  $\mu_{cf}$  and  $\sigma_{cf}$ .

### 3.2 Entry, Investment, and Exit

Since an incumbent's value  $V(k, s)$  is weakly increasing in the idiosyncratic productivity shock  $s$  and the conditional distribution  $H(s'|q)$  is decreasing in the signal  $q$ , the value of entering  $V_e(q)$  is a strictly increasing function of the signal. In turn, this means that there will be a threshold for  $q$ , call it  $q^*$ , such that prospective entrants will enter if and only if they receive a draw greater than or equal to  $q^*$ .

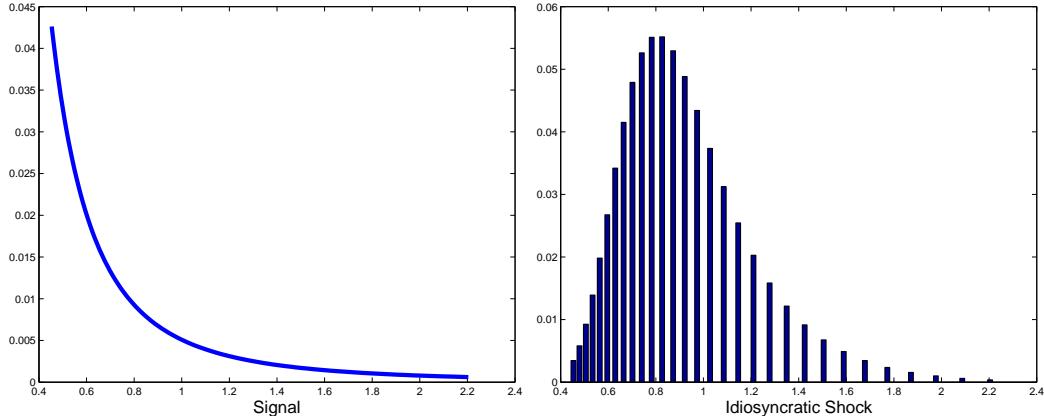


Figure 2: Signal's distribution (left) and productivity distribution of actual entrants.

The mass of actual entrants with productivity less than or equal to any  $\bar{s} \geq 0$  will be  $M \int_{q^*}^{\bar{s}} dH(s|q)dQ(q)$ . See the right panel of Figure 2.

Let  $k'(q)$  denote the optimal entrants' capital choice conditional on having received a signal  $q$ . For every  $\bar{k} \geq k'(q^*)$  and for all  $\bar{s} \geq 0$ , the portion of entrants characterized by pairs  $(s, k)$  such that  $s \leq \bar{s}$  and  $k \leq \bar{k}$  will be  $M \int_0^{\bar{s}} \int_{q^*}^{q^{-1}(\bar{k})} dQ(q)dH(s|q)$ , where  $q^{-1}$  is the inverse of the mapping between signal and capital choice.

Our treatment of the entry problem is different from that in [Hoppenhayn \(1992\)](#). There,

prospective entrants are identical. Selection upon entry is due to the fact that firms that paid the entry cost start operating only if their first productivity draw is greater than the exit threshold. In our framework, prospective entrants are heterogeneous. Some obtain a greater signal than others and therefore face better short-term prospects. As a consequence, they start with greater capital and are likely to hire more employees.

Investment is modeled as in the standard neoclassical framework with non-convex adjustment costs. See for example Khan and Thomas (2008). The fixed cost of investment gives rise to policy functions of the S-s type, which in turn are responsible for the periods of inaction that are characteristic of all firm- and plant-level datasets.

Firms exit whenever they draw an operating cost  $c_f$  such that the value of exiting  $V_x(k)$  is greater than the value of continuing operations, i.e.  $\tilde{V}(k, s) - c_f$ . Figure 3 depicts the probability of survival as a function of the state variables. For given capital, the probability of exiting is decreasing in the idiosyncratic productivity. This is not surprising, given that the value of exiting does not depend on  $s$ , while the continuation payoff is obviously greater, the larger is productivity. Survival chances also improve with capital, for given level of productivity. Both the value of continuing and exiting are increasing and concave in capital, but simple inspection reveals that the former grows faster with  $k$ .

Differently from Hopenhayn (1992), where firms exit with certainty when productivity falls below a certain threshold and continue with certainty otherwise, in our framework all firms survive with positive probability and survival is smoothly increasing in both capital and productivity. If we let the distribution of the operating cost collapse to a singleton, the exit policy would be characterized by a decreasing schedule, call it  $\underline{s}(k)$ , such that a firm equipped with capital  $k$  will exit if and only if its productivity is lower than  $\underline{s}(k)$ .

### 3.3 Calibration

Our parameter values are listed in Table 1. One period is assumed to be one year. Consistent with most macroeconomic studies, we assume that  $R = 1.04$ ,  $\delta = 0.1$ , and  $\alpha = 0.3$ . We set  $\theta$ , which governs returns to scale, equal to 0.8. This value is on the lower end of the range of estimates recovered by Basu and Fernald (1997) using industry-level data. Using plant-level data, Lee (2005) finds that returns to scale in manufacturing vary from 0.828 to 0.91, depending on the estimator.

The remaining parameters are chosen in such a way that a number of statistics computed using a panel of simulated data are close to their empirical counterparts. Unless indicated otherwise, the simulated data is drawn from the stationary distribution. We list

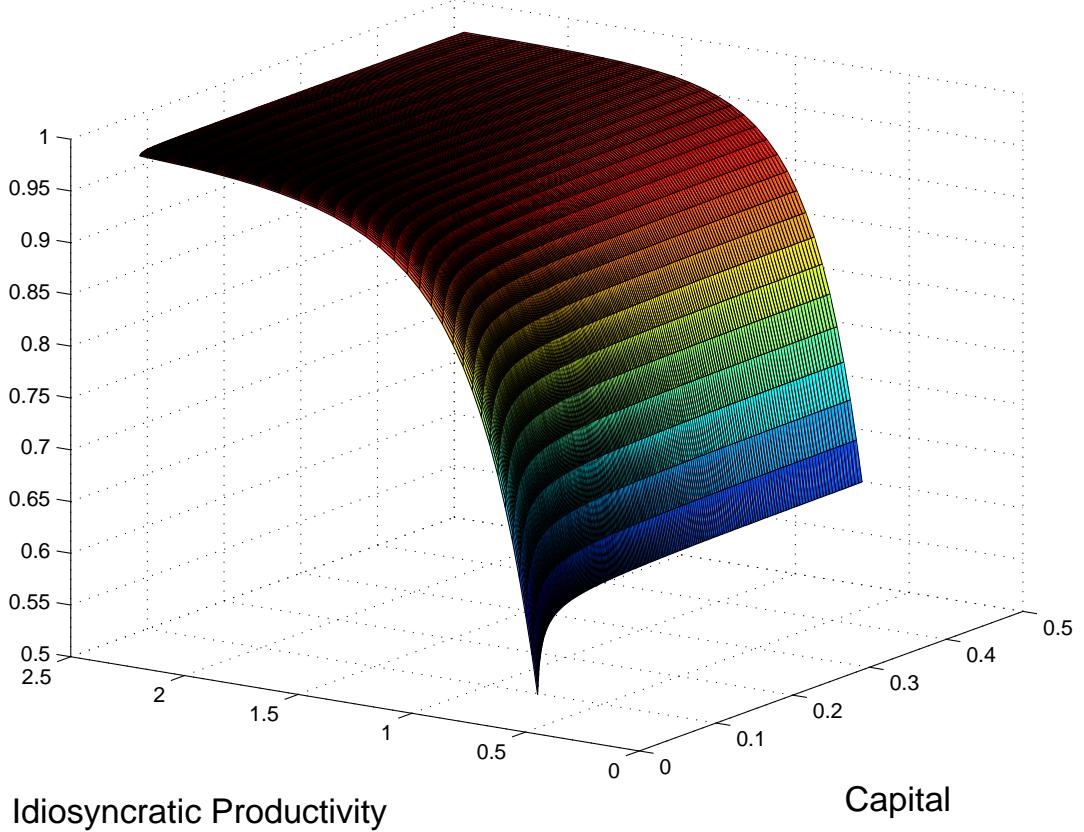


Figure 3: Conditional probability of survival.

the simulated moments and their empirical counterparts in Table 2.

Because of non-linearities, it is not possible to match parameters to moments. However, the mechanics of the model clearly indicates which are the key parameters for each set of moments. What follows is a summary description of the algorithm we used to select parameter values.

First, notice that there are uncountably many pairs  $(M, \gamma)$  which yield stationary equilibria identical to each other except for the volume of operating firms (and entrants). All the statistics of interest for our study are the same across all such allocations. To see why this is the case, start from a given stationary equilibrium and consider raising the labor supply elasticity  $\gamma$ . The original equilibrium wage will now elicit a greater supply of labor. Then, simply find the new, greater value for  $M$ , such that the aggregate demand for labor in stationary equilibrium equals supply at the original wage.

We borrow the value of  $\gamma$  from the calibration of the general case with  $\sigma_z > 0$ , to be illustrated in Section 4. In that model, the property just pointed out does not hold true and the parameter is pinned down by the volatility of employment growth. Given this

choice, we set  $M$  in such a way that the equilibrium wage is 3.0. This value is clearly arbitrary. However – it will be clear at the end of this section – it involves no loss of generality.

Description	Symbol	Value
Capital share	$\alpha$	0.3
Span of control	$\theta$	0.8
Depreciation rate	$\delta$	0.1
Interest rate	$R$	1.04
Labor supply elasticity	$\gamma$	2.0
Mass of potential entrants	$M$	1,766.29
Persistence idiosync. shock	$\rho_s$	0.55
Variance idiosync. shock	$\sigma_s$	0.22
Operating cost – mean parameter	$\mu_{c_f}$	-5.63872
Operating cost – var parameter	$\sigma_{c_f}$	0.90277
Fixed cost of investment	$c_0$	0.00011
Variable cost of investment	$c_1$	0.03141
Pareto exponent	$\xi$	2.69
Entry cost	$c_e$	0.005347

Table 1: Parameter Values.

The parameters of the process driving the idiosyncratic shock, along with those governing the adjustment costs, were chosen to match the mean and standard deviation of the investment rate, the autocorrelation of investment, and the rate of inaction, i.e. the fraction of periods in which the investment rate is less than 1%. The targets are the moments computed by [Cooper and Haltiwanger \(2006\)](#) using a balanced panel from the LRD from 1972 to 1988. Because of selection in entry and exit, such statistics are likely to be biased estimates of the population moments. To ensure consistency between the simulated statistics and their empirical counterparts, we follow [Cao \(2007\)](#) and compute our moments on balanced panels extracted from the simulated series.

A simpler version of the neoclassical investment model with lognormal disturbances – one without investment adjustment costs – has the interesting properties that (i) the mean investment rate is a simple non linear function of the parameters  $\rho_s$  and  $\sigma_s$  and that (ii) the standard deviation of the investment rate is a simple non-linear function of the mean. It follows that in that framework, mean and standard deviation do not identify the pair  $\{\rho_s, \sigma_s\}$ . While these properties do not hold exact in our model, inspection reveals that similar restrictions exist. We proceed to set  $\sigma_s = 0.22$ , a value consistent with available

estimates of the standard deviation of the innovation to idiosyncratic productivity shocks,<sup>6</sup> and set  $\rho_s$  to minimize a weighted average of the distances of mean and standard deviation of the investment rate from their targets.<sup>7</sup>

The remaining parameters are set to match the average entry rate and the relative size of entrants and exiters with respect to survivors, respectively. The definitions of these ratios are those of [Dunne, Roberts, and Samuelson \(1988\)](#) and the targets are the statistics obtained by [Lee and Mukoyama \(2015\)](#) using the LRD. The mean of the operating cost has the largest effect on the exit rate, which in stationary equilibrium must equal the entry rate. The variance of the operating cost pins down the relative size of exiters.

Finally, the pair  $\{\xi, c_e\}$  determines the relative size of entrants. For simplicity, we set  $c_e$  equal to the mean operating cost, and then pick the value of the Pareto exponent to hit the target.

Statistic	Model	Data
Mean investment rate	0.153	0.122
Std. Dev. investment rate	0.325	0.337
Investment autocorrelation	0.059	0.058
Inaction rate	0.067	0.081
Entry rate	0.062	0.062
Entrants' relative size	0.58	0.60
Exiters' relative size	0.47	0.49

Table 2: Calibration Targets.

We claimed above that our arbitrary choice of  $M$  implies no loss of generality. How would the stationary equilibrium look like if, say, we chose a greater value? The wage rate would be higher and therefore all firm sizes would decline. Exit would be more appealing, which in turn would imply higher exit (and entry) rate. However, firms dynamics will be exactly the same, as the investment process is invariant to the wage rate. It follows that by revising our choices for the parameters governing entry and exit, we can generate an economy that differs from ours only in its scale. All the moments of interest will be the same, and so will be the model's implications for entry, growth, and survival.

---

<sup>6</sup>See [Haltiwanger \(2011\)](#) and [Castro, Clementi, and Lee \(2015\)](#).

<sup>7</sup>[Castro, Clementi, and Lee \(2015\)](#) provide direct sector-specific estimates of  $\rho_s$  for 140 3-digit SIC manufacturing industries. Their estimates range from 0.3 to 0.7, with a weighted mean of 0.43, a value also close to what reported by [Abraham and White \(2006\)](#). Exploiting information on 11 seven-digit SIC product categories, [Foster, Haltiwanger, and Syverson \(2008\)](#) come up with an estimate of 0.8.

### 3.4 Firm Dynamics

In this section, we describe the model's implications for firm growth and survival, and compare them with the empirical evidence. Unless otherwise noted, size is measured by employment. Productivity is simply the value of the random variable  $s_t$ . In Section 5.4, we argue that the appropriate empirical counterpart to the latter is revenue total factor productivity (TFPR), as defined in [Foster, Haltiwanger, and Syverson \(2008\)](#).

The left panel of Figure 4 illustrates the unconditional relation between exit hazard rate and age. Consistent with [Dunne, Roberts, and Samuelson \(1989\)](#) and all other studies we are aware of, the exit hazard rate decreases with age. This is the case because on average entrants are less productive than incumbents. As a cohort ages – see the right panel – the survivors' productivity and value increase, leading to lower exit rates.

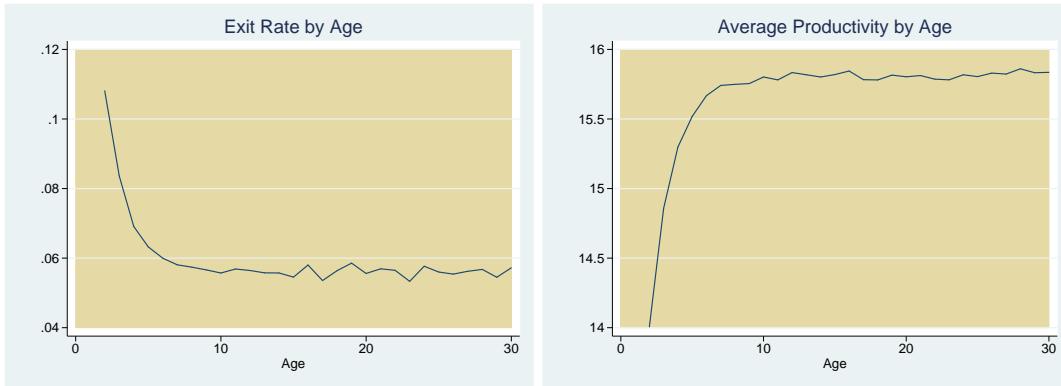


Figure 4: The exit hazard rate

A similar mechanism is at work in [Hopenhayn \(1992\)](#). In his framework, however, there exists a size threshold such that the exit rate is 100% for smaller firms and identically zero for larger firms. This feature is at odds with the evidence.<sup>8</sup> In our model, firms with the same employment are characterized by different combinations of  $(k, s)$  and therefore have different continuation values. Those with relatively low capital and relatively high productivity are less likely to exit.

[Dunne, Roberts, and Samuelson \(1989\)](#) also found that in the US manufacturing sector, establishment growth is unconditionally negatively correlated with both age and size – a finding that was confirmed for a variety of sectors and countries.<sup>9</sup> [Evans \(1987\)](#) and [Hall \(1987\)](#) found evidence that firm growth declines with size even when we condition on age, and viceversa.

<sup>8</sup>See [Dunne, Roberts, and Samuelson \(1989\)](#) and [Lee and Mukoyama \(2015\)](#).

<sup>9</sup>See [Coad \(2009\)](#) for a survey of the literature.

Hopenhayn (1992) is consistent with these facts, with the exception of the conditional correlation between growth and age. In his model, idiosyncratic productivity is a sufficient statistics for firm size and growth. Conditional on age, smaller firms grow faster because the stochastic process is mean-reverting. However, firms of the same size behave identically, regardless of their age. The model generates the right unconditional relation between age and growth, simply because age and size are positively correlated in the stationary distribution. When controlling for size, age is uncorrelated with growth.

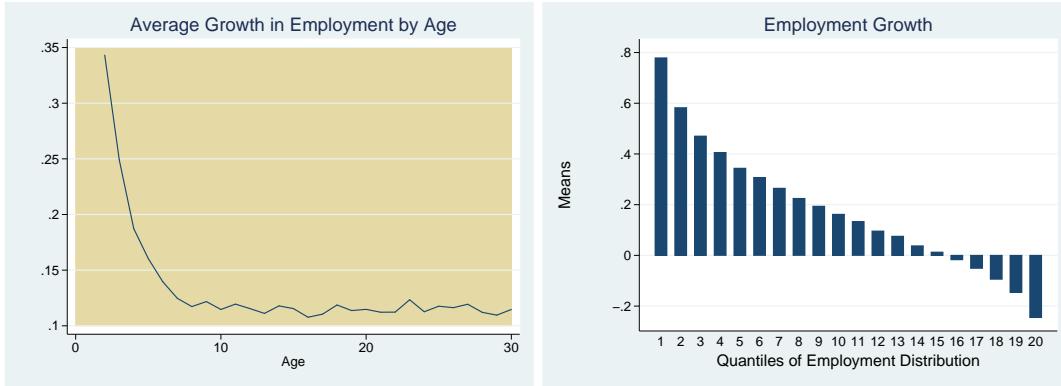


Figure 5: Unconditional Relationship between Growth, Age, and Size.

Our version of the model is consistent with all the facts about growth listed above. Figure 5 illustrates the unconditional correlations. Recall that the state variables are productivity and capital. Conditional on age, employment growth declines with size because larger firms tend to have higher productivity levels. Given that productivity is mean-reverting, their growth rates will be lower.

Now consider all the firms with the same employment. Since adjustment costs prevent the instantaneous adjustment of capital to the first-best size implied by productivity, some firms will be characterized by a relatively low capital and high shock, and others by a relatively high capital and low shock. The former will grow faster, because investment and employment are catching up with the optimal size dictated by productivity. The latter will shrink, as the scale of production is adjusted to the new, lower level of productivity.

The conditional negative association between age and size arises because, on average, firms with relatively high  $k$  and low  $s$  – which are shrinking – will be older than firms with low  $k$  and high  $s$  – which are growing. For shrinking firms, productivity must have declined. For this to happen, they must have had the time to grow in the first place. On average, they will be older than those that share the same size, but are growing instead.

The model is consistent with the evidence on firm growth even when we measure size

with capital rather than employment. Conditional on age, capital is negatively correlated with growth for the same reason that employment is. It is still the case that larger firms have higher productivity on average. Another mechanism contributes to generating the right conditional correlation between growth and size. Because of investment adjustment costs, same-productivity firms have different capital stocks. It turns out that on average the larger ones have seen their productivity decline, while the smaller ones have seen their efficiency rise. The former are in the process of shrinking, while the latter are growing.

We just argued that firms with the same capital will have different productivity levels. For given capital, firms with higher shocks are growing, while firms with lower shocks are shrinking. Once again, the negative conditional correlation between growth and age follows from the observation that, on average, firms with higher shocks are younger.

It is worth emphasizing that, no matter the definition of size, the conditional relation between age and growth is driven by relatively young firms. Age matters for growth even when conditioning on size, because it is (conditionally) negatively associated with productivity. To our knowledge, only two other papers present models that are consistent with this fact. The mechanism at work in D’Erasco (2011) is similar to ours. Cooley and Quadrini (2001) obtain the result in a version of Hopenhayn (1992)’s model with financial frictions and exogenous exit.

The left panel of Figure 6 shows the firm size distribution that obtains in stationary equilibrium. Noticeably, it displays positive skewness. The right panel illustrates the evolution of a cohort size distribution over time. Skewness declines as the cohort ages. Both of these features are consistent with the evidence gathered by Cabral and Mata (2003) from a comprehensive data set of Portuguese manufacturing firms.

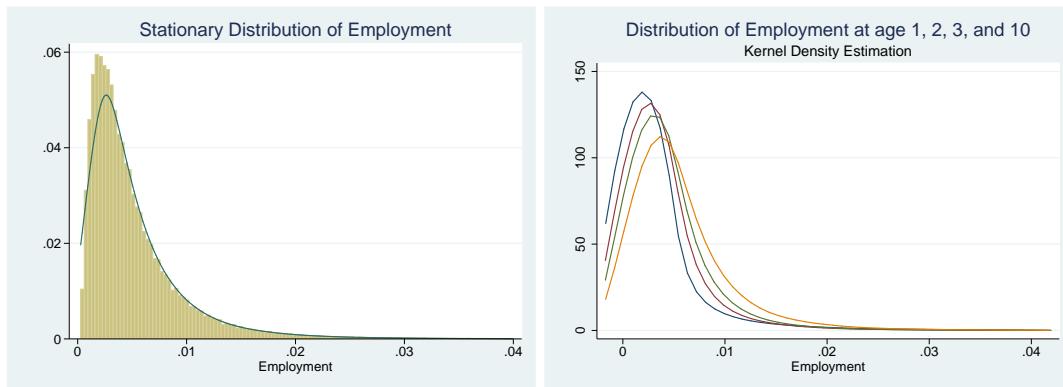


Figure 6: Evolution of a Cohort’s Size Distribution

## 4 Aggregate Fluctuations – Mechanics

We now move to the scenario with aggregate fluctuations. In order to formulate their choices, firms need to forecast next period's wage. The labor market clearing condition implies that the equilibrium wage at time  $t$  satisfies the following restriction:

$$\log w_t = \frac{\log[(1-\alpha)\theta z_t]}{1 + \gamma[1 - (1-\alpha)\theta]} + \frac{1 - (1-\alpha)\theta}{1 + \gamma[1 - (1-\alpha)\theta]} \Omega_t, \quad (1)$$

with  $\Omega_t = \log \left[ \int (sk^{\alpha\theta})^{\frac{1}{1-(1-\alpha)\theta}} d\Gamma_t(k, s) \right]$ . The log-wage is an affine function of the logarithm of aggregate productivity and of a moment of the distribution.

Unfortunately, the dynamics of  $\Omega_t$  depends on the evolution of  $\Gamma_t$ . It follows that the vector of state variables  $\lambda_t$  consists of the distribution  $\Gamma_t$  and the aggregate shock  $z_t$ . Faced with the formidable task of approximating an infinitely-dimensional object, we follow Krusell and Smith (1998) and conjecture that  $\Omega_{t+1}$  is an affine function of  $\Omega_t$  and  $\log z_{t+1}$ . Then, (1) implies that the law of motion for the equilibrium wage is

$$\log w_{t+1} = \beta_0 + \beta_1 \log w_t + \beta_2 \log z_{t+1} + \beta_3 \log z_t + \varepsilon_{t+1}, \quad (2)$$

where  $\varepsilon_{t+1}$  is a zero-mean random variable.

When computing the numerical approximation of the equilibrium allocation, we will impose that firms form expectations about the evolution of the wage assuming that (2) holds true. This means that the aggregate state variables reduce to the pair  $(w_t, z_t)$ . The parameters  $\{\beta_0, \beta_1, \beta_2, \beta_3\}$  will be set equal to the values that maximize the accuracy of the prediction rule. The definition of accuracy and its assessment are discussed in Section 4.2. The algorithm is described in detail in Appendix A.

### 4.1 Calibration

With respect to the stationary case, we need to calibrate three more parameters. These are  $\rho_z$  and  $\sigma_z$ , which shape the dynamics of aggregate productivity, and the labor supply elasticity  $\gamma$ . We set them in order to match the standard deviation and auto-correlation of output growth, as well as the standard deviation of employment growth.

The targets for the first two are the standard deviation and autocorrelation of the growth in non-farm private value added from 1947 to 2008, from the US Bureau of Economic Analysis. The third target is the standard deviation of employment growth, also in the non-farm private sector and for the same period, from the Bureau of Labor Statistics.

Our parameter values are listed in Table 3. Our choice for  $\gamma$  is close to recent estimates of the aggregate labor supply elasticity. For example, Rogerson and Wallenius (2009)

Description	Symbol	Value
Labor supply elasticity	$\gamma$	2.0
Persist. aggregate shock	$\rho_z$	0.685
Std. Dev. aggregate shock	$\sigma_z$	0.0163

Table 3: Parameter Values.

conclude that for micro elasticities ranging from 0.05 to 1.25, the corresponding macro elasticities are in the range of 2.25 to 3.0. [Fiorito and Zanella \(2012\)](#) estimate a micro elasticity of 0.1 and macro elasticities in the range 1.1-1.7. At the quarterly frequency, our values for  $\rho_z$  and  $\sigma_z$  imply an autocorrelation of about 0.91 and a standard deviation of about 0.008. In comparison, [Cooley and Prescott \(1995\)](#) set the two parameters at 0.95 and 0.007, respectively.

Statistic	Model	Data
Standard deviation output growth	0.032	0.032
Autocorrelation output growth	0.069	0.063
Std. dev. employment growth (rel. to output growth)	0.656	0.667

Table 4: Additional Calibration Targets.

## 4.2 The Forecasting Rule

The forecasting rule for the equilibrium wage turns out to be

$$\log(w_{t+1}) = 0.38385 + 0.65105 \log(w_t) + 0.53075 \log(z_{t+1}) - 0.21508 \log(z_t) + \varepsilon_{t+1}.$$

The wage is persistent and mean-reverting. A positive aggregate shock increases the demand for labor from both incumbents and entrants. This is why  $\beta_2$ , the coefficient of  $\log(z_{t+1})$ , is estimated to be positive. The coefficient of  $\log(z_t) - \beta_3$  – is negative because the larger the aggregate shock in the previous period, the smaller is going to be the expected increment in aggregate productivity, and therefore the lower the wage increase.

In the literature with heterogeneous agents and aggregate risk it has become standard to evaluate the accuracy of the forecasting rule by assessing the  $R^2$  of the regression, which in our case is 0.9971. However, as pointed out by [Den Haan \(2010\)](#), this choice is questionable on at least three grounds. To start with, the  $R^2$  considers predictions made conditional on wages generated by the true law of motion. In this sense, it only assesses the accuracy of one-period ahead forecasts. Second, the  $R^2$  is a measure of

average accuracy. In the numerical methods literature, it is standard to report maximum errors instead. Last, but not least, the  $R^2$  scales the error term by the variance of the dependent variable. The problem here is that it is often not clear what the appropriate scaling is. The root mean squared error (0.00103 in our case) does not suffer from the latter shortcoming, but is affected by the first two.

Here we follow Den Haan (2010)' suggestion to assess the accuracy of our forecasting rule by calculating the maximum discrepancy (in absolute value) between the actual wage and the wage generated by the rule *without* updating. That is, we compute the maximum pointwise difference between the sequence of actual market-clearing wages and that generated by our rule, when next period's predicted wage is conditional on last period's prediction for the current wage rather than the market clearing wage. The value of that statistic over 24,500 simulations is 0.765% of the market clearing wage.

The frequency distribution of percentage forecasting errors is illustrated in the left panel in Figure 7. The absolute value of the forecasting error is greater than 0.5% of the clearing wage in only 1.3% of the observations. The right panel is a scatter plot of equilibrium wages and their respective forecasts. It shows that the points are aligned along the  $45^\circ$  line and that forecasting errors are small with respect to the variation in equilibrium wages. More diagnostics is reported in the Appendix.

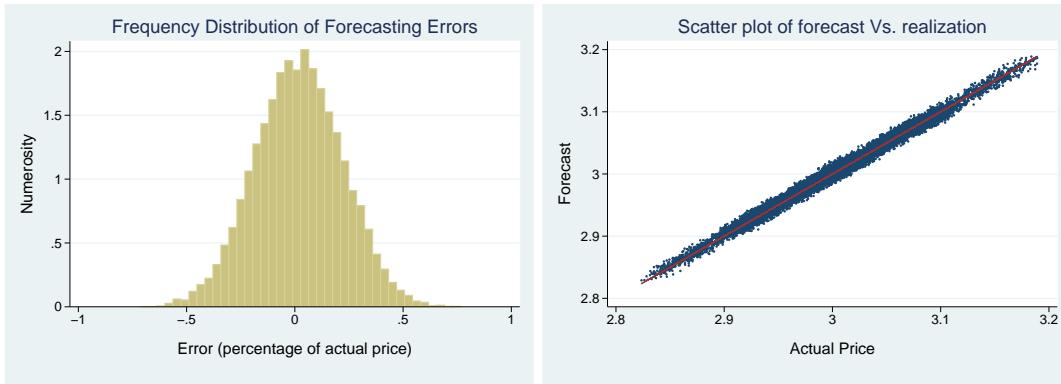


Figure 7: Accuracy of the Forecasting Rule.

### 4.3 Entry and Exit

Recall that in the stationary equilibrium analyzed in Section 3, the solution to the entry problem consists of a time-invariant threshold on the signal space and a policy function for investment. Here the threshold will be time-varying, and will depend on the wage and on the aggregate productivity realization.

The value of an incumbent is strictly increasing in aggregate productivity and strictly decreasing in the wage, while the cost of entry is constant. It follows that the entry threshold will be greater, the lower is aggregate productivity and the greater is the wage. See the left panel in Figure 8 for an illustration.

Everything else equal, a rise in aggregate productivity (or a decline in the wage) will lead to an increase in the number of entrants. Given that the distribution of idiosyncratic shocks is stochastically increasing in the value of the signal, such a rise will also lead to a decline in entrants' average idiosyncratic efficiency.<sup>10</sup>

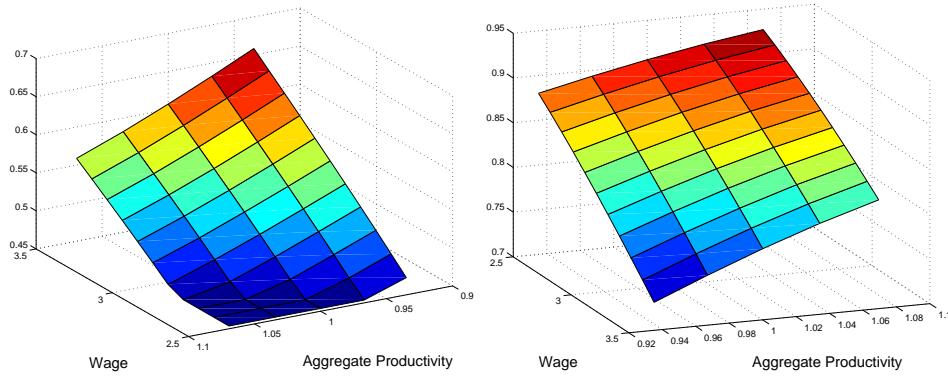


Figure 8: Left: Entry Threshold on the Signal Space. Right: Survival Probability.

The conditional probability of survival is also time-varying. Since the value recovered upon exit is not a function of aggregate variables, survival chances are greater the higher is aggregate productivity and the lower is the wage. The right panel of Figure 8 illustrates how the probability of continuing to operate changes with the two aggregate state variables, for a given pair of individual states  $\{k, s\}$ .

Everything equal, a positive shock to aggregate productivity leads to a decline in the number of exiting firms. Since the distribution of the idiosyncratic shock is invariant over time and across firms, the average idiosyncratic efficiency of exiters will also decline.

The equilibrium dynamics of the entry and exit margins will obviously depend on the response of the wage to innovations in aggregate productivity. The wage and aggregate shock will be strongly positively correlated. The milder the response of the wage to positive aggregate productivity shocks, the higher the likelihood that periods of high aggregate TFP will be characterized by high entry volumes, high survival rates, and

---

<sup>10</sup> For high levels of aggregate productivity and low wages the schedule on the left panel of figure 8 is flat. In that portion of the aggregate state space, the signal threshold coincides with its lower bound and all potential entrants find it optimal to start producing. It turns out, however, that this portion of the state space does not belong to the ergodic set.

relatively low productivity of entrants and exiters.

## 5 Aggregate Fluctuations – Results

### 5.1 Cyclical Behavior of Entry and Exit

Table 5 reports the raw correlations of entry rate, exit rate, and the size of entrants and exiters (relative to incumbents) with industry output. Consistent with the evidence presented by [Campbell \(1998\)](#), the entry rate is pro-cyclical and the exit rate is counter-cyclical.

Entry Rate	Exit Rate	Entrants' Size	Exiters' Size
0.402	-0.779	-0.725	-0.892

Table 5: Correlations with industry output.

Interestingly, [Campbell \(1998\)](#) also provides evidence on the correlations between entry rate, exit rate, and future and lagged output growth. He finds that the correlation of entry with lagged output growth is greater than the contemporaneous correlation and that the exit rate is positively correlated with future output growth.

Our model is consistent with both features. The correlation of entry rate with contemporaneous output growth is 0.0039, while that with one-period lagged output growth is 0.8359. The reason is that the entry decision is taken contingent on the information available one period before the start of operations. The correlation coefficient between exit rate and one- and two-period ahead output growth are 0.13 and 0.26, respectively. Periods of low exit tend to be periods of high output. Given the mean-reverting nature of aggregate productivity, on average such periods are followed by times of low output growth.

Analyzing data from the LRD, [Lee and Mukoyama \(2015\)](#) find that selection at entry is quantitatively very important. Relative to incumbents, entering plants tend to be larger during recessions. This is the case, according to their evidence, because entrants' average idiosyncratic productivity is counter-cyclical. Our model shares these features of the data. When the common productivity component is low, only firms with a relatively high level of idiosyncratic productivity find it worthwhile to enter.

The banking literature also found evidence in support of the claim that aggregate conditions have an impact on selection at entry. A number of papers, among which [Cetorelli \(2014\)](#), find that when credit market conditions are relatively favorable, entering

firms are less productive on average.

In our model, the relative size of exiters is also higher during recessions. A drop in the common productivity component leads to a lower value of all incumbents. It follows that the marginal exiter will have a higher value of the idiosyncratic productivity component.

## 5.2 Impulse Responses

In order to gain some more intuition about the model's dynamics, we now describe the impulse response functions to a hike in  $z_t$ . We initialize the system by assuming that the distribution of firms is equal to the point-wise mean of distributions on the ergodic set. The common productivity component is set at its mean value. At time  $t = 1$ , we impose that the exogenous aggregate productivity component rises by about  $3.5\% - 1.5$  standard deviations – and we compute the evolution of the size distribution over the next 25 periods. We repeat this experiment for 3,000 times and depict the averages of selected variables in Figures 9 and 10.

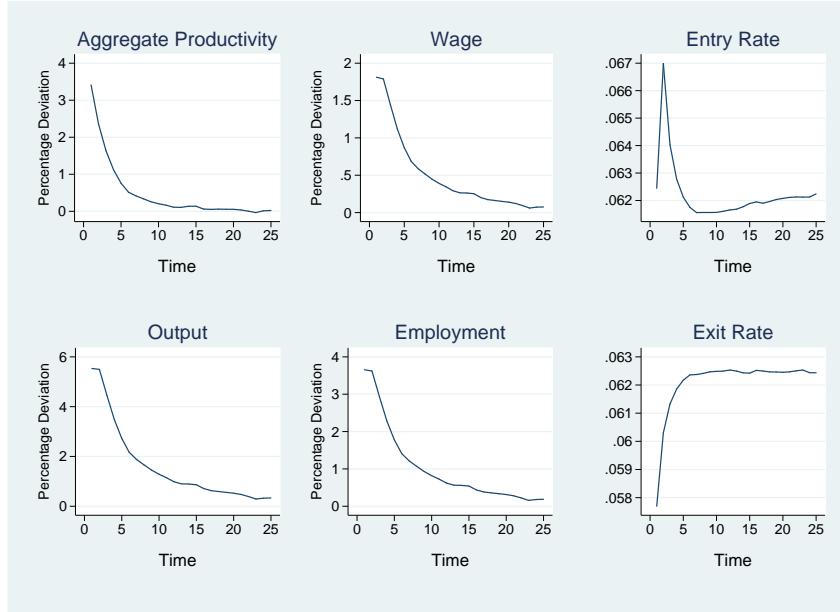


Figure 9: Response to a positive productivity shock.

The left-most panel on the top of Figure 9 reports the percentage deviation of the common productivity component from its unconditional mean. Not surprisingly, output, the wage rate and employment display qualitatively similar dynamics.

Recall that, because of our timing assumptions, entry is governed by past aggregate conditions and exit occurs after production has taken place. It follows that the contemporaneous response of output, employment, and wage to the positive productivity shock

is entirely due to the expansion in hiring by incumbents.

The compositional effects of the aggregate shock appear in the next period. The exit rate declines, while the entry rate rises.<sup>11</sup> As a result – see Figure 10 – the number of operating firms also rises. It peaks in period 5, when the exit volume, which is rising back towards its unconditional mean, overcomes the declining entry volume.

Figure 10 also illustrates the dynamics of selection in entry and exit. The average size and idiosyncratic productivity of entrants (relative to incumbents) declines and then converges back to its unconditional mean. This is the case because, on balance, the hike in aggregate productivity and the ensuing increase in the wage imply a rise in the value of incumbency. It follows that the average idiosyncratic productivity of the pool of entrants declines.

The average size and idiosyncratic productivity of exiters (relative to non-exiters) also declines. The improvement in aggregate factors implies an increase in the value of continuing operations. As a result, the marginal firms, i.e. those indifferent between exiting and staying, decline in quality. Since entry increases while exit declines, the cross-sectional mean of idiosyncratic productivity decreases.

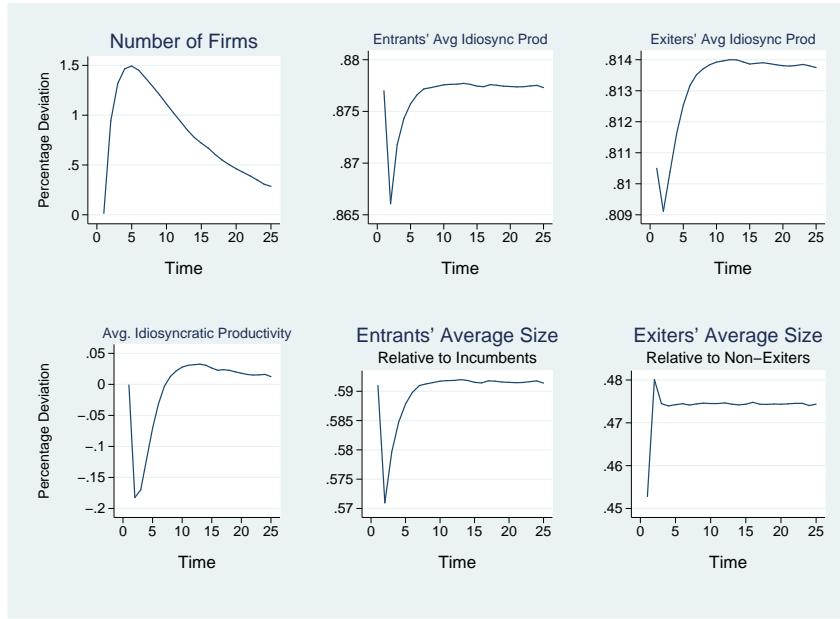


Figure 10: Response to a positive productivity shock.

Notice that the convergence of entry rate and exit rate to their respective unconditional

<sup>11</sup>Following [Dunne, Roberts, and Samuelson \(1988\)](#), the period- $t$  exit rate is the ratio of firms exiting between period  $t$  and  $t + 1$  and the number of operating firms in time  $t$ . This statistic drops in the period where the shocks hits, but the effect of exit on the distribution only appears in the next period.

means is not monotone. The exit rate overshoots its long-run value. The entry rate undershoots. Similarly, for all the statistics illustrated in Figure 10, except for the mass of operating firms. This happens because the wage decays at a lower pace than aggregate productivity.

A few periods after the positive shock hits, the exogenous component of aggregate productivity is back close to its unconditional mean, but the wage is not. The recent evolution of the number of operating firms and their distribution over idiosyncratic productivity is responsible for keeping aggregate labor demand higher than simply warranted by  $z_t$  and the customary capital accumulation channel. We will elaborate at length on this key point in Section 6. With a relatively high wage and relatively low  $z_t$ , the selection effect changes sign. Entry falls below its long-run value, while exit is higher than that.

### 5.3 Alternative assumptions on the discount factor

As assumed in Section 2 and maintained throughout the paper, firms discount future cash flows at a constant rate. This would not be the case in a straightforward extension of our framework to general equilibrium with risk-averse households. With complete markets, for example, shareholders would unanimously support firms' choices taken discounting future values by means of the intertemporal marginal rate of substitution.

In general, such rate will be time-varying and structurally linked to aggregate forcing variables, such as aggregate productivity. It is appropriate, then, to ask how the predictions of our model would change if we dispensed with our assumption of a constant discount factor.

The answer to this question hinges upon the covariance between the discount factor and aggregate productivity, which in our model is intimately linked to the cyclicity of the discount factor. Everything else equal, amending our model by assuming a pro-cyclical discount factor is going to amplify the impact of entry. This is the case, because in times of high aggregate productivity, characterized by relatively high entry, a higher-than-average discount factor will make entry even more appealing. As a result, the covariance between entry rate and output would increase. With a counter-cyclical discount factor, the opposite will be true.

The empirical evidence provides strong support for the pro-cyclical scenario. No-arbitrage implies that the conditional expectation of the stochastic discount factor equals the reciprocal of the risk-free rate. No matter the time horizon and the technique employed to tease out inflation expectations, the real risk-free rate is always found to be counter-cyclical, indicating that the discount factor is indeed pro-cyclical. See [Beaudry and Guay](#)

(1996) and Cooper and Willis (2014) among others.

One may wonder, then, under what assumptions extending our analysis to general equilibrium would yield a pro-cyclical discount factor. Based on the work of Beaudry and Guay (1996) and Winberry (2014), we argue that assuming habit formation in consumption would be sufficient to generate this pattern.

Beaudry and Guay (1996) were the first to point out that the prototypical real business cycle model of King, Plosser, and Rebelo (1988), which features time-separable additive CRRA preferences over consumption sequences, produces a counter-factually pro-cyclical risk-free rate (a counter-cyclical conditional expected discount factor), unless one is willing to assume unrealistically low values for the autocorrelation of aggregate productivity.

The intuition is straightforward. A positive and persistent technology shock drives up investment demand via an increase in the marginal product of capital and simultaneously increases consumption demand through an increase in permanent income. These two shifts in aggregate demand outstrip the hike in aggregate supply, causing current goods to become relatively scarce and the interest rate to rise.

Beaudry and Guay (1996) showed that introducing habit formation in consumption can overturn the prediction for the cyclicity of the risk-free rate, by muting the contemporaneous response of consumption to innovations in productivity. More recently, Winberry (2014) showed that the same holds true in the context of a state-of-the-art model which would be the natural extension of our framework to general equilibrium, except for the fact that he does not model entry and exit.

#### 5.4 Alternative assumptions on idiosyncratic shocks

In our setup, idiosyncratic shocks only affect technical efficiency. In reality, however, plants also face random shifts in input supply and product demand schedules, which lead to changes in input and product prices, respectively.

The main implications of our theory do not hinge upon the actual nature of plant-level shocks. In particular, they will still hold in setups with horizontal product differentiation where idiosyncratic demand shocks play a major role, as long as such innovations are isomorphic to our own disturbances. In other words, the appropriate definition of plant-level productivity in our model is that of Revenue Total Factor Productivity (TFPR).

Exploiting data from the five Census of Manufacturing in the period 1977–1997, Foster, Haltiwanger, and Syverson (2008, 2015) conclude that in the case of ten seven-digit industries, the dynamics of TFPR was mostly the result of idiosyncratic demand shocks. Entrants displayed about the same technical efficiency as incumbents, but faced much

lower demand schedules, and therefore charged lower prices. Over time, conditional on survival, demand shifted outwards, leading to higher prices and greater TFPR.

To rationalize these findings, Foster, Haltiwanger, and Syverson (2015) posit a stochastic process for idiosyncratic demand shocks which leads to a TFPR dynamics very similar to ours. Therefore we conjecture that our results would not change if we amended our model to allow for idiosyncratic random variation in demand.

## 6 The relevance of entry and exit for aggregate dynamics

In this section we show that allowing for entry and exit enhances the model’s internal amplification and propagation mechanism. The rise in entry and decline in exit prompted by a hike in the exogenous component of aggregate productivity leads to a rise in the number of operating firms and an increase in the skewness of their distribution. In other words, it leads to more firms characterized by small size but high conditional growth rates. Their evolution over time keeps the demand for labor high, above and beyond the level warranted by  $z_t$  and the customary capital accumulation channel.

The role of entrants in shaping aggregate productivity dynamics also emerges from a simple decomposition exercise, that reveals the contribution of net entry to aggregate productivity growth to be positive. As in the data, the entrants’ role grows larger over time.

Finally, we describe our model’s implications for the cyclicalities of selected cross-sectional moments.

### 6.1 Amplification and Propagation

We begin by comparing our model – the benchmark, from now on – to a simplified version of it that features neither exit nor entry. At every point in time, there is a constant mass of firms, set equal to the unconditional average number of incumbents that obtains along the benchmark’s equilibrium path. Figure 11 depicts the impulse responses of industry output to a positive aggregate shock in the two scenarios, respectively.

We argued above that in the benchmark case the contemporaneous response of output is entirely due to incumbents. Therefore it is not surprising that the period-1 percentage deviation from the unconditional mean is the same across the two economies. Starting in period 2, however, output is greater in the benchmark economy. This is the result of changes in the mass of active firms as well as their distribution over the state space.

To clarify this key insight, we appeal briefly to a simplified version of our model, one in which labor is the only input. Let the individual production function be  $y_t = s_t z_t l_t^\alpha$ .

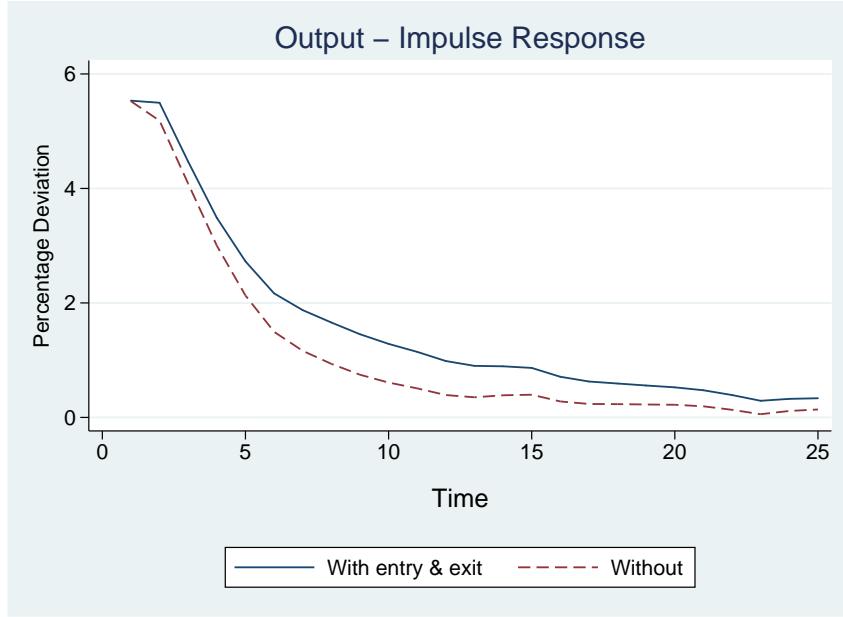


Figure 11: The effect of entry & exit on output dynamics.

Simple calculations allow us to write total output as  $Y_t = z_t [\int \Gamma_t(s) s^{\frac{1}{1-\alpha}} ds]^{1-\alpha} L_t^\alpha$ , where  $L_t$  is aggregate employment and  $\Gamma_t$  is the measure of firms over idiosyncratic productivity, the only individual state variable. Letting  $N_t \equiv \int \Gamma_t(s) ds$  and  $\widehat{\Gamma}_t \equiv \Gamma_t/N_t$ , we can then rewrite

$$Y_t = z_t \left[ \int \widehat{\Gamma}_t(s) s^{\frac{1}{1-\alpha}} ds \right]^{1-\alpha} N_t^{1-\alpha} L_t^\alpha. \quad (3)$$

Aggregate total factor productivity is the product of  $z_t$  and an *allocative efficiency* factor which reflects the distribution of firms over idiosyncratic productivity. Aggregate inputs are labor and the number of active firms.

In the case without entry or exit,  $\widehat{\Gamma}_t$  converges to the density associated with the stationary distribution of  $s_t$ , i.e. the density of a log-normal with parameters 0 and  $\frac{\sigma_s^2}{1-\rho_s^2}$ , respectively. Since  $N_t$  is also constant in that scenario,  $Y_t$  derives all of its time-series properties from the process driving  $z_t$ .

In the scenario with entry and exit which is the focus of this paper instead, both  $\widehat{\Gamma}_t$  and  $N_t$  are time-varying and their respective dynamics contribute to shaping the marginal product of labor, along with  $z_t$ . Via its impact on the entry and exit margins, a positive innovation to the exogenous component of aggregate productivity leads the number of firms to rise in the next period. The density  $\widehat{\Gamma}_t$  is also affected, reflecting in particular the larger-than-usual influx of low-productivity firms. At later times,  $N_t$  and  $\widehat{\Gamma}_t$  are driven by the dynamics of entry and conditional exit rates, as well as the process for idiosyncratic

productivity growth implied by  $s_t$ . What's important is that the evolution of  $N_t$  and  $\hat{\Gamma}_t$  implies that  $Y_t$  mean-reverts at a lower pace than implied by  $z_t$ .

Our digression sheds new light on Figure 11. The two economies share  $z_t$  and the standard propagation role of the physical capital stock. In the case with entry and exit, however, the evolution in the mass of firms and their distribution keeps the marginal product of labor higher. Refer once more to Figure 10. The number of firms peaks only at  $t = 5$  and slowly reverts to its unconditional mean. At birth, firms born in the aftermath of the shock are very small and therefore account for a rather small fraction of total output. However, conditional on survival, they grow in efficiency and size faster than incumbents. This process takes place at the same time in which the aggregate productivity component regresses towards its unconditional mean. As a result, aggregate output falls at a slower pace.

To the extent that, in reality, the actual birth of new firms – net of the associated purchase of physical capital – requires purposeful investment, one can legitimately think of  $N_t$  as a form of intangible capital, albeit one with time-varying depreciation.<sup>12</sup> In this sense, the impact of entry and exit on persistence is not entirely surprising, as productivity shocks are now propagated by investment in two reproducible factors, with an overall elasticity in production larger than usual. An interesting question is whether our model is indeed isomorphic to a more standard setup with tangible as well intangible investment, as the one considered in McGrattan and Prescott (2010) among others.

The simple scenario without physical capital briefly considered above indicates with clarity that this is not the case. Entry and exit, together with data-conforming firm dynamics, are also responsible for a time-varying *allocative efficiency* factor, whose evolution contributes to shaping the response of aggregate labor demand to shocks. This mechanism is absent in standard aggregative models of accumulation. In the next subsection, we will consider further evidence on the importance of such factor in the data and in our model.

Before plunging into that, however, we wish to assess the magnitude of the amplification induced by firm demographics. Figure 12 depicts percentage deviations of cumulative output from the unconditional mean, in the scenarios with and without entry and exit, respectively.<sup>13</sup> Consistent with our discussion above, the difference is small on impact, but grows over time. After 10 years, cumulative output with entry and exit is 2.97% higher

---

<sup>12</sup>Notice also that in a general equilibrium extension of our model, proper accounting of the resources expended in starting new firms would lead to an upward revision of aggregate investment and value added.

<sup>13</sup>If  $t$  denotes the number of periods since the shock hit, aggregate output is  $Y_s$ , and the unconditional mean aggregate output is  $\bar{Y}$ , Figure 12 reproduces the mean of  $[\sum_{s=1}^t Y_s]/t\bar{Y}$ , calculated over 3,000 simulations.

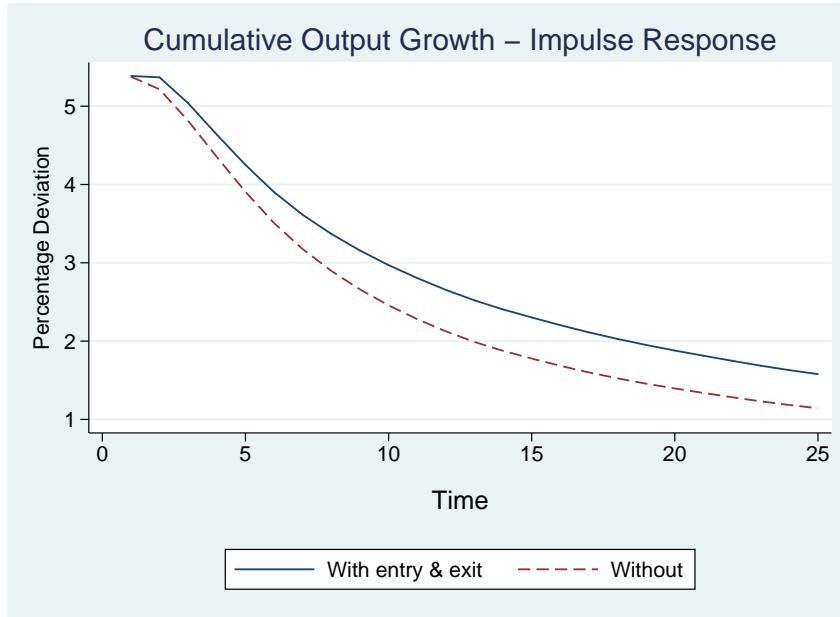


Figure 12: Cumulative Output Growth.

than the unconditional mean value. Without entry or exit, it is only 2.46% higher. In other words, on average entry and exit are responsible for almost one fifth of above-trend output growth over the 10 years following the shock.

Given our modeling choices and parametric assumptions, this is likely to be only a lower bound for the impact of entry and exit. In our framework, plants of different cohorts share the same technology. If, along the lines of [Campbell \(1998\)](#), we let entrants adopt the most recent and productive vintage of machinery, the impact of novel cohorts would be magnified accordingly.

Notice also that our calibration of entry and exit is extremely conservative. Bound by the requirement that all moments targeted in the calibration of firm-level parameters are calculated on the same data, we elected to set the unconditional mean entry rate at 6.2%. However, data from the US Census Bureau's Business Dynamics Statistics Database show that the economy-wide mean plant entry rate over the period 1977-2011 was more than twice as large, at 12.72%, with a standard deviation of 0.91%.<sup>14</sup>

We also study the dynamics of the Solow residual in the two economies. The residual is computed by assuming an aggregate production function of the Cobb–Douglas form with capital share equal to 0.3. Extrapolating from what we learnt above in the context of a simplified version of our model, in the benchmark the Solow residual is shaped

---

<sup>14</sup>Entry rates declined steadily over this period. The standard deviation refers to the linearly de-trended series.

by the common productivity component  $z_t$ , the distribution of firms over idiosyncratic productivity, and the mass of active firms.

To quantify the impact of entry and exit on persistence, we conducted the experiment of setting the parameters in the economy without entry and exit in such a way that it generates the same values of the target moments as the benchmark economy. To generate the target autocorrelation for output growth, we had to set  $\rho_z = 0.775$ , much higher than the value of 0.685 assumed in Section 4. The difference is explained by the fact in the case without entry or exit, both the mass of firms and their distribution over idiosyncratic productivity are constant.

This simple exercise hints that trying to infer information about the process generating factor-neutral exogenous technical change using a model without entry or exit will give the wrong answer. Such model will inevitably interpret improvements in aggregate efficiency that result from fast productivity growth in young firms, as well as the reallocation of output shares towards them, as changes in the exogenous aggregate component.

## 6.2 Productivity Decomposition

We have just argued that a key reason why in our model aggregate output is more persistent than implied by the forcing variable  $z_t$ , is that entrants' productivity grows faster than incumbents', drawing a reallocation of market share towards them. Strong evidence in support of the latter claim comes from the literature that exploits longitudinal establishment data to study the determinants of aggregate productivity growth.

After thoroughly reviewing that literature, Foster, Haltiwanger, and Krizan (2001) conclude that “*studies that focus on high-frequency variation tend to find a small contribution of net entry to aggregate productivity growth while studies over a longer horizon find a large role for net entry.*” They go on to add that “*Part of this is virtually by construction... Nevertheless, ... The gap between productivity of entering and exiting plants also increases in the horizon over which the changes are measured since a longer horizon yields greater differential from selection and learning effects.*” That is, conditional on survival entrants' productivity grows faster.

A productivity decomposition exercise carried out on simulated data generated by our model yields results which are qualitatively consistent with the evidence illustrated by Foster, Haltiwanger, and Krizan (2001). On average, the contribution of net entry to productivity growth is positive, as entering firms tend to be more productive than the exiters they replace. Its magnitude is small when the interval between observations is one period (equivalent to one year), but it increases with the time between observations. In

part, this is due to the mere fact that the output share accounted for by entrants is larger, the longer the horizon over which changes are measured. However, it is also due to the fact that entrants grow in size and productivity at a faster pace than incumbents.

Define total factor productivity as the weighted sum of firm–level Solow residuals, where the weights are the output shares. Let  $\mathcal{C}_t$  denote the collection of plants active in both periods  $t - k$  and  $t$ . The set  $\mathcal{E}_t$  includes the plants that entered between the two dates and are still active at time  $t$ . In  $\mathcal{X}_{t-k}$  are the firms that were active at time  $t - k$ , but exited before time  $t$ .

Following [Haltiwanger \(1997\)](#), the growth in TFP can be decomposed into five components, corresponding to the addenda in equation (4). They are known as (i) the within component, which measures the changes in productivity for continuing plants, (ii) the between–plant portion, which reflects the change in output shares across continuing plants, (iii) a covariance component, and finally (iv) entry and (v) exit components.<sup>15</sup>

$$\begin{aligned} \Delta \log(TFP_t) = & \sum_{i \in \mathcal{C}_t} \phi_{i,t-k} \Delta \log(TFP_{it}) + \sum_{i \in \mathcal{C}_t} [\log(TFP_{i,t-k}) - \log(TFP_{t-k})] \Delta \phi_{it} + \\ & \sum_{i \in \mathcal{C}_t} \Delta \log(TFP_{it}) \Delta \phi_{it} + \sum_{i \in \mathcal{E}_t} [\log(TFP_{it}) - \log(TFP_{t-k})] \phi_{it} - \\ & \sum_{i \in \mathcal{X}_{t-k}} [\log(TFP_{i,t-k}) - \log(TFP_{t-k})] \phi_{i,t-k} \end{aligned} \quad (4)$$

Table 6 reports the results that obtain when we set  $k$  equal to 1 and 5, respectively. In the last column, labeled Net Entry, we report the difference between the entry and exit components. Recall that in our model the unconditional mean of aggregate productivity growth is identically zero.

$k$	Within	Between	Covariance	Entry	Exit	Net Entry
1	-8.7778	-4.1623	12.4366	-0.2964	-0.8011	0.5047
5	-15.1437	-13.3884	27.5833	-0.7556	-1.7104	0.9548

Table 6: Productivity Decomposition (percentages).

The between and within components are necessarily negative, because of mean reversion in the process driving idiosyncratic productivity. Larger firms, which tend to be more productive, shrink on average. Smaller firms, on the contrary, tend to grow. The covariance component is positive, because firms that become more productive also increase in size.

---

<sup>15</sup>With  $\phi_{it}$  and  $TFP_{it}$  we denote firm  $i$ 's output share and measured total factor productivity at time  $t$ , respectively.  $TFP_t$  is the weighted average total factor productivity across all firms active at time  $t$ .

On average, both the entry and exit contributions are negative. This reflects the fact that both entrants and exiters are less productive than average. However, entrants tend to be more productive than exiters. The contribution of net entry to productivity growth is positive regardless of the horizon.

What's most relevant for our analysis is that for  $k = 5$  the contribution of net entry is about twice that for  $k = 1$ . In part, this is due to the fact that the share of output produced by entrants increases with  $k$ . However, this cannot be the whole story. The contribution of entry is roughly  $-0.3\%$  for  $k = 1$  and goes to  $-0.76\%$  for  $k = 5$ . If entrants' productivity did not grow faster than average, the contribution of entry for the  $k = 5$  horizon would be much smaller.

### 6.3 Cyclical Variation of Cross-sectional Moments

Starting with [Eisfeldt and Rampini \(2006\)](#), a number of papers have assessed the cyclical behavior of the cross-sectional distribution of productivity. Particular interest has been paid to the evolution of the second moment.

Exploiting the German dataset USTAN and data from the US Census of Manufacturing, respectively, [Bachman and Bayer \(2013\)](#) and [Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry \(2014\)](#) conclude that the standard deviation of idiosyncratic TFP growth is countercyclical. [Kehrig \(2015\)](#), also analyzing US Census data, reports that a measure of dispersion of the level of TFP is countercyclical as well.

In spite of our assumption that the stochastic process governing idiosyncratic productivity is constant across time and across firms, the equilibrium allocation of our model features non-trivial dynamics of the cross-sectional distribution of productivity. The reason, as it is by now clear, is that the net entry rate, as well as selection in entry and exit, are time-varying.

The impulse response functions discussed in section 5 hint that when industry output is high, the idiosyncratic productivity of both entrants and exiters is relatively low. Since the entry-rate is pro-cyclical while the exit-rate is counter-cyclical, it follows that the first moment of the idiosyncratic productivity distribution is counter-cyclical.

Another robust implication of the model is that skewness is pro-cyclical. In Section 3 we showed that, consistent with the evidence, the productivity distribution is positively skewed. During expansions, both entrants and exiters are less productive than average. Since the net entry rate is pro-cyclical, skewness increases.

This feature is illustrated in Figure 13, which plots the difference between the (average) fraction of firms associated with each level of idiosyncratic shock in expansion

and in recession, respectively. When industry output is high, there are relatively more low-productivity firms and relatively less high-productivity firms. The right tail of the distribution is thinner.

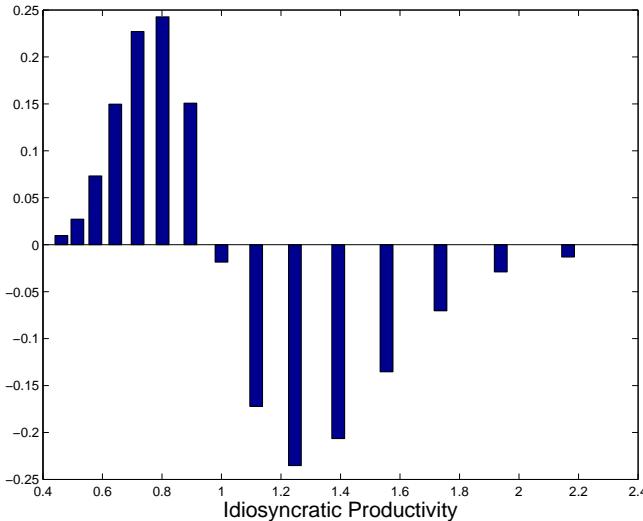


Figure 13: Change in the Cross-Sectional Distribution.

Unfortunately the prediction for the cyclicity of the second moment is not robust. Our simulations show that the sign of the correlation between second moment and industry output changes with relatively minor changes in parameter values.

## 7 Accounting for the Slow Recovery from the Great Recession

The 2007–2009 recession stands out not only for its depth, but also for the very slow recovery that followed it – in particular for employment. According to the Current Employment Statistics gathered by the Bureau of Labor Statistics (BLS), it took longer than six years for total non-farm employment to regain the level it had reached in January 2008.

Another feature of that recession – perhaps less known but equally striking – is the unusually large and protracted drop in the number of establishments. Figure 14 depicts the evolution of the number of private-sector plants according to the BLS’ Business Employment Dynamics (BED) and the Bureau of Census’ Business Dynamics Statistics (BDS), respectively.<sup>16</sup>

---

<sup>16</sup>A similar dynamics emerges when considering firm- rather than plant-level data, also available from the BDS.

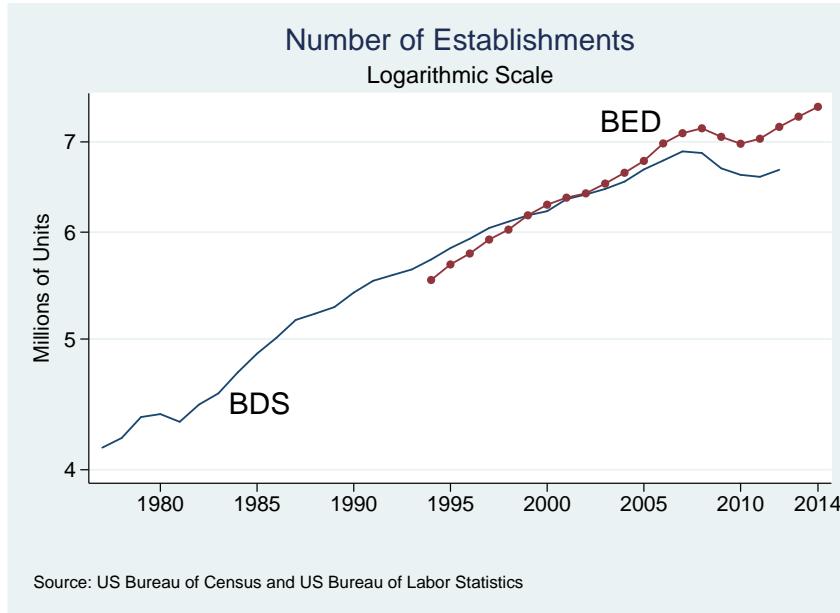


Figure 14: Number of private-sector establishments.

To a large extent, the drop in the number of productive units was due to the poor performance of a few cohorts born during the great recession. The middle panel in Figure 15 substantiates this claim by depicting the fraction of total establishments accounted for by the cohorts born in 2006 through 2009 in the first four years since birth, respectively. The data is from the BDS and the shaded area denotes the 2-standard deviation interval around the time-series mean of each statistic, allowing for comparison with all prior cohorts.<sup>17</sup>

The 2006 cohort – included as a benchmark – appears to be average, accounting for about 12% of establishments at birth and 8% three years later. The 2009 cohort, on the other end, is more than two standard deviations away from the mean generation, accounting for only about 9% of operating establishments at birth and less than 6% by the time it turned of age 3.

---

<sup>17</sup>With reference to Figure 15, let  $n_{ts}$  denote the number of establishments born in year  $t$  and still in operation at  $t+s$ . Let also  $e_{ts}$  denote their total employment at  $t+s$ . For the cohorts born in  $t = 2006, 2007, 2008, 2009$ , the employment share  $s$  years after birth – for  $s = 0, 1, 2, 3$  – is  $e_{ts}/E_{t+s}$ , where  $E_{t+s}$  is total employment in year  $t+s$ . The cohort's extensive margin is  $n_{ts}/N_{t+s}$ , where  $N_{t+s}$  is the total number of establishments at  $t+s$ . The intensive margin is  $(e_{ts}/n_{ts})/(E_{t+s}/N_{t+s})$ . In accordance with what reported by Jarmin and Miranda (2002), we find that the number of entrants (age 0 establishments) is significantly larger during census years (years ending in 2 and 7), after controlling for a linear trend. To attenuate the implied bias in the calculation of the extensive margin, we filter the size of each entering cohort using a dummy variable that takes value of zero during census years and one otherwise. The size of the year- $t$  cohort  $s$  years since birth equals the number of establishments at entry multiplied by the survival rate between time  $t$  and  $t+s$ .

The right panel shows that, conditional on survival, plants belonging to all cohorts considered were fairly average. At birth, their mean employment was slightly more than half the average employment of incumbent plants – a figure close to the time-series average for that statistic – and also grew over time at an average pace.

The lesson we draw is that the 2009 cohort, and to a slightly lesser extent their 2008 peers, appear to have suffered dramatically on the extensive margin – their numerosity – but not nearly as much on the intensive margin – the average plant size relative to incumbents.

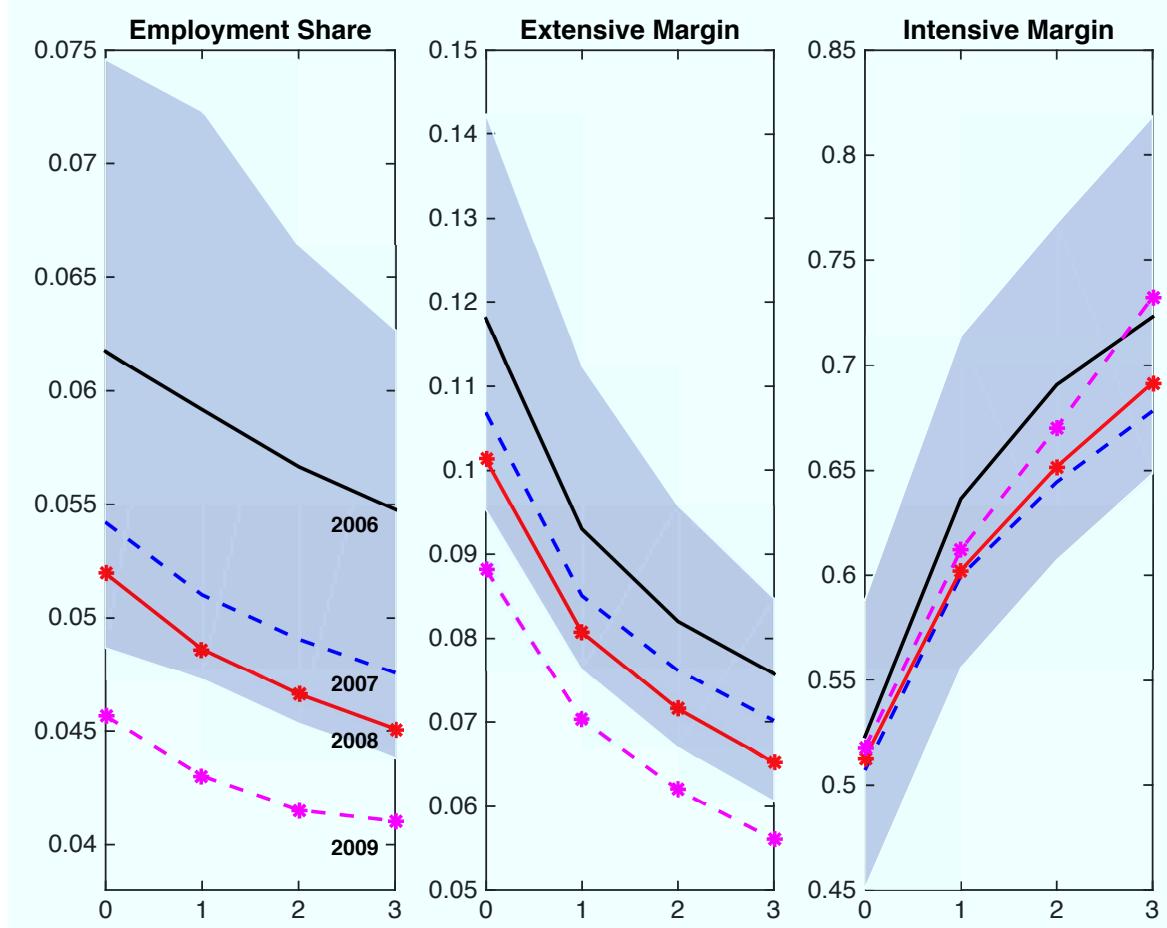


Figure 15: Young Firms during the Great Recession.

The product of the quantities depicted in the middle and right panels yields the fraction of total employment accounted for by a given cohort, the statistic better suited to quantify its contribution to the dynamics of aggregate employment. We illustrate it in the left panel. Comparing the employment share of the 2009 cohort with the mean employment share across all cohorts reveals that, modulo equilibrium effects, the lower numerosity

(with respect to average) of the 2009 cohort accounts for about 1.8% of private-sector employment from birth through 2012.

Thus plant demographics may have had a sizeable and persistent effect on aggregate outcomes. Our model illustrates how. As an immediate corollary of the results illustrated in Section 5, a negative shock to aggregate productivity will generate a persistent decline in the number of establishments, which in turn will slow down the recovery of aggregate employment. In the remainder of this section, we will assess the model’s ability to generate dynamics that are also quantitatively close to those just documented.

Recall that our calibration strategy, outlined in Section 3.3, relies on comprehensive micro data for a specific sample of establishments, i.e. manufacturing plants over the period 1977-1997. According to the BDS data, which include all private-sector establishments, the mean annual entry rate is roughly twice as large as our target of 6.2%. Since recalibrating the model to allow a greater role to the entry (and exit) margin would introduce a logical inconsistency, we decide against it. We retain our parameter values and evaluate the result of our assessment as a lower bound on the role played by plant-level demographics in shaping the recovery from the great recession.

Refer to Figure 16. The solid lines illustrate the evolution of selected variables, following a negative 4% time-0 mean-reverting shock to the exogenous component of aggregate productivity. The magnitude of the impulse was chosen to generate a drop in aggregate employment of about 4%, a value which is half-way between the employment losses with respect to trend that result by applying to BED data the log-linear and Hodrick-Prescott filter, respectively.

The fall in entry and the hike in exit lead to a persistent decline in the number of operating establishments, which bottoms out three years after the initial impulse. In the middle panel of Figure 17, we can appreciate that the fraction of establishments accounted for by the generation born at time 1 – the solid line – is lower than the fraction accounted for by the typical cohort born in average aggregate conditions, at the same age (the dash-dotted line). The contribution of that generation towards employment is also lower than the typical cohort’s. See the left panel.<sup>18</sup>

Notice also that, consistent with the evidence presented above, members of the generation born at  $t = 1$  that survive do as well – relative to incumbents – as their peers born in normal times. See the right panel in Figure 17. The totality of the effect of plant-level demographics comes from the extensive margin – the drop in cohort’s size.

---

<sup>18</sup>Comparing the left panels of Figures 15 and 17 reveals that, unlike the data, our model predicts that a cohort’s contribution to employment be increasing early in life.

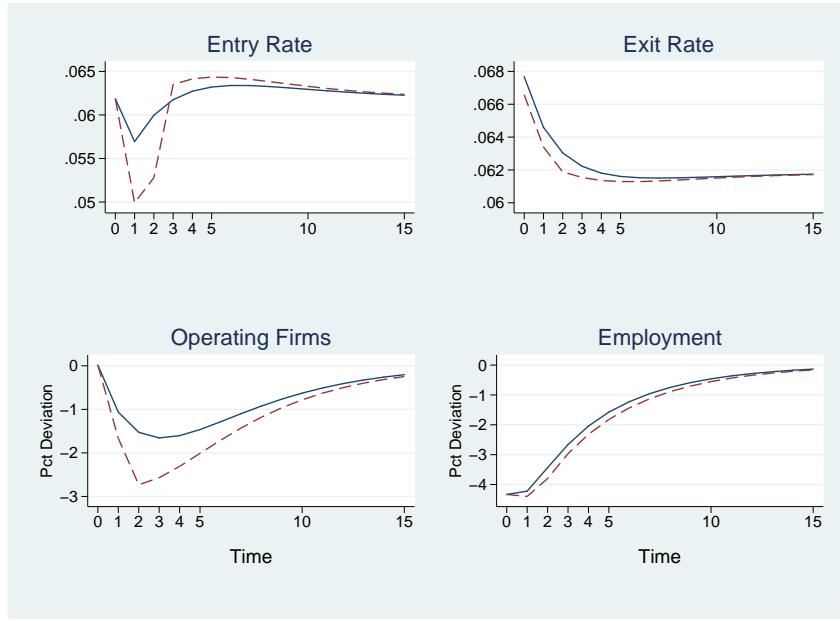


Figure 16: Solid line: TFP shock only. Dashed line: TFP plus entry-specific shock.

The shock to  $z_t$  delivers a maximum cumulative drop in the number of establishments just a little higher than 1.5%, short of what found in the data. According to the BED, the number of establishments reached its trough in 2010, at a level 2.6% smaller than its previous peak, in 2008.

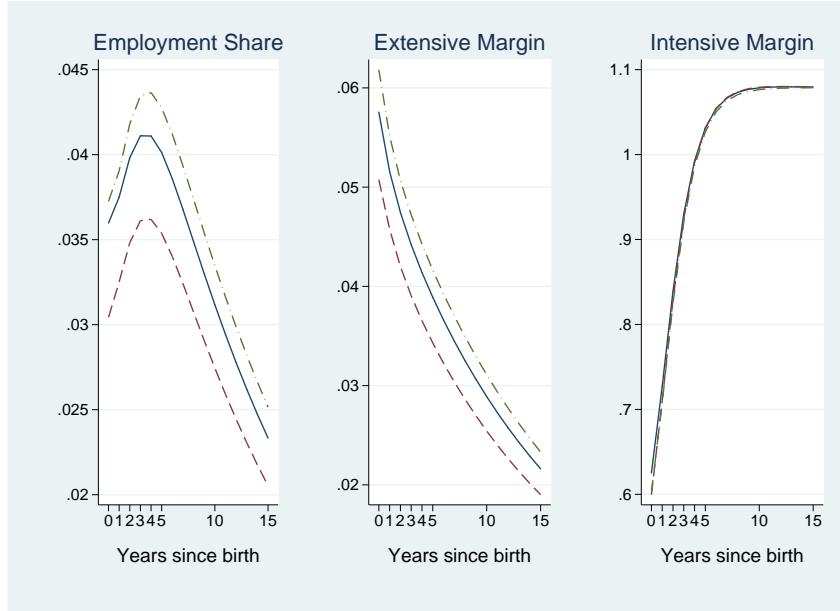


Figure 17: Solid: TFP shock only. Dashed: TFP plus entry-specific shock. Dot-dashed: Average cohort.

Read through the lenses of our model, the data tells us us that at the outset of the last recession the entry margin suffered to an extent that cannot have been simply the result of an aggregate shock. Thus we decide to investigate the incremental effect of an exogenous disturbance exclusively affecting entrants. We assume that for  $t = 0, 1$ , potential entrants' mean conditional idiosyncratic productivity decreases by 3% and then reverts to its long-run value. This choice was informed by the goal of obtaining a time path for the number of establishments closer to the data.<sup>19</sup>

The resulting dynamics are illustrated by the dashed lines in Figures 16 and 17. With respect to the scenario with the TFP shock only, the decline in entry is larger, while the increase in exit is less pronounced. The latter feature results from a lower wage rate, which makes exit less appealing.

The contribution of the  $t = 1$  cohort to both the establishment and employment counts is lower at all times: The shock sustained by that generation is permanent. Interestingly, once again, all action is on the extensive margin. Conditional on survival, young firms compare to incumbents similarly to their peers born in normal times.<sup>20</sup>

Notice that the size of the overall impact of the entry-specific shock on aggregate quantities is muted by the lower wage, which boosts hiring by incumbents. However, its impact on persistence of aggregate series is substantial. The half-life of employment rises from five to six years. Interestingly, although entrants' idiosyncratic productivity is back to normal by  $t = 2$ , the same date in which the establishment count reaches its trough, the marginal contribution of the entry shock to the fall in aggregate employment increases monotonically until  $t = 12$ , when it equals 17% of the total.

## 8 Conclusion

This paper provides a framework to study the dynamics of the cross-section of firms and its implications for aggregate dynamics. When calibrated to match a set of moments of the investment process, our model delivers implications for firm dynamics and for the cyclicity of entry and exit that are consistent with the evidence.

The survival rate increases with size. The growth rate of employment is decreasing

---

<sup>19</sup>Tracing out the effects of the entry-specific shock in the full-blown model characterized in Sections 4 and 6 would have been computationally prohibitively demanding. Figures 16 and 17 depict the perfect-foresight transitional dynamics resulting from hitting the stationary model analyzed in Section 3 with unexpected (zero-measure) shocks.

<sup>20</sup>The top left panel of Figure 16 also shows that in the scenario with both shocks the entry rate reverts faster to its long-run value. The reason is that the lower entry volumes at  $t = 1, 2$  moderate wage growth. When the mean conditional productivity of potential entry returns to normal, entry becomes substantially more appealing with respect to the case of TFP shock only.

with size and age, both unconditionally and conditionally. The size distribution of firms is skewed to the right. When tracking the size distribution over the life of a cohort, the skewness declines with age. The entry rate is positively correlated with current and lagged output growth. The exit rate is negatively correlated with output growth and positively associated with future growth.

Carefully modeling firm-level dynamics turns out to be key. The pro-cyclicality of entry and the positive association between age and firm growth deliver amplification and propagation of aggregate shocks in a plain-vanilla competitive framework.

A positive shock to aggregate productivity leads to an increase in entry. Consistent with the empirical evidence, entrants are smaller than incumbents. The skewness of the distribution of firms over idiosyncratic productivity increases. As the exogenous component of aggregate productivity declines towards its unconditional mean, the new entrants that survive grow in productivity and size. That is, the distribution of idiosyncratic productivity improves. As a result, the response of output is stronger and more persistent than in an environment that abstracts from entry and exit.

Our numerical experiments reveal that on average, entry and exit account for about one fifth of the above-trend growth experienced by our economy over the 10 years following a 1.5 standard deviation innovation to aggregate productivity. As an alternative metric of the impact of entry and exit on aggregate dynamics, we assessed the effect on persistence. For a version of our model without entry or exit to generate a data-conforming persistence of output, the first-order autocorrelation of aggregate productivity shocks must be 0.775. In the benchmark scenario with entry and exit, it needs only be 0.685.

Last, but not least, we show that according to our model there is a clear causal link between the exceptionally large drop in establishments during the great recession and the painfully low speed of the recovery from it. Whatever its exact nature, the adverse event that initiated the recession had a particularly strong effect on the 2008 and 2009 cohorts, severely reducing the ranks of small but high-conditional-growth plants and thereby suppressing the growth in aggregate labor demand for years to come.

In spite of the wealth of detail and descriptive realism achieved by the model, our framework could be extended in a variety of dimensions. In particular, we would like to gauge the quantitative effect of our mechanism when imbedded in a genuine general equilibrium framework. Relaxing our assumption that firms of different cohorts share the same technology is also of interest. Assuming, as it appears to be the case in reality, that entrants are more likely to adopt more recent vintages of capital, is likely to further enhance the amplification and propagation mechanism uncovered here.

## A Numerical Approximation

Our algorithm consists of the following steps.

1. Guess values for the parameters of the wage forecasting rule  $\hat{\beta} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3\}$ ;
2. Approximate the value function of the incumbent firm;
3. Simulate the economy for  $T$  periods, starting from an arbitrary initial condition  $(z_0, \Gamma_0)$ ;
4. Obtain a new guess for  $\hat{\beta}$  by running regression (2) over the time series  $\{w_t, z_t\}_{t=S+1}^T$ , where  $S$  is the number of observations to be scrapped because the dynamical system has not reached its ergodic set yet;
5. If the new guess for  $\hat{\beta}$  is close to the previous one, stop. If not, go back to step 2.

### A.1 Approximation of the Value Function

The incumbent's value function is approximated by value function iteration.

1. Start by defining grids for the state variables  $w, z, k, s$ . Denote them as  $\Psi_w, \Psi_z, \Psi_k$ , and  $\Psi_s$ , respectively. The wage grid is equally spaced and centered around the equilibrium wage of the stationary economy. The capital grid is constructed following the method suggested by [McGrattan \(1999\)](#). The grids and transition matrices for the two shocks are constructed following [Tauchen \(1986\)](#). For all pairs  $(s, s')$  such that  $s, s' \in \Psi_s$ , let  $H(s'|s)$  denote the probability that next period's idiosyncratic shock equals  $s'$ , conditional on today's being  $s$ . For all  $(z, z')$  such that  $z, z' \in \Psi_z$ , let also  $G(z'|z)$  denote the probability that next period's aggregate shock equals  $z'$ , conditional on today's being  $z$ .
2. For all triplets  $(w, z, z')$  on the grid, the forecasting rule yields a wage forecast for the next period (tildes denote elements not on the grid):

$$\log(\tilde{w}') = \hat{\beta}_0 + \hat{\beta}_1 \log(w) + \hat{\beta}_2 \log(z') + \hat{\beta}_3 \log(z).$$

In general,  $\tilde{w}'$  will not belong to the grid of wages. There will be contiguous grid points  $(w_i, w_{i+1})$  such that  $w_i \leq \tilde{w}' \leq w_{i+1}$ . Now let  $J(w_i|w, z, z') = 1 - \frac{\tilde{w}' - w_i}{w_{i+1} - w_i}$ ,  $J(w_{i+1}|w, z, z') = \frac{\tilde{w}' - w_i}{w_{i+1} - w_i}$ , and  $J(w_j|w, z, z') = 0$  for all  $j$  such that  $j \neq i$  and  $j \neq i+1$ . This allows us to evaluate the value function for values of the wage which are off the grid by linear interpolation;

3. For all grid elements  $(w, z, k, s)$ , guess values for the value function  $V_0(w, z, k, s)$ ;
4. The revised guess of the value function,  $V_1(w, z, k, s)$ , is determined as follows:

$$V_1(w, z, k, s) = \pi(w, z, k, s) + \Pr[c_f > c_f^*(w, z, k, s)] V_x(k) \\ + \Pr[c_f \leq c_f^*(w, z, k, s)] [\tilde{V}(w, z, k, s) - E[c_f | c_f \leq c_f^*(w, z, k, s)]]$$

subject to

$$\pi(w, z, k, s) = \frac{1 - (1 - \alpha)\theta}{(1 - \alpha)\theta} w^{-\frac{\theta(1-\alpha)}{1-\theta(1-\alpha)}} [(1 - \alpha)\theta s z k^{\alpha\theta}]^{\frac{1}{1-\theta(1-\alpha)}}, \\ V_x(k) = k(1 - \delta) - g[-k(1 - \delta), k], \\ \tilde{V}(w, z, k, s) = \max_{k' \in \Psi_k} \left\{ -x - c_0 k \chi - c_1 \left(\frac{x}{k}\right)^2 k \right. \\ \left. + \frac{1}{R} \sum_j \sum_i \sum_n V_0(w_i, z_j, k', s_n) H(s_n | s) J(w_i | w, z, z_j) G(z_j | z) \right\}, \\ x = k' - k(1 - \delta), \\ \chi = 1 \text{ if } k' \neq k \text{ and } \chi = 0 \text{ otherwise,}$$

where  $c_f^*(w, z, k, s)$  is the value of the fixed operating cost that makes the firm indifferent between the choice of exiting and obtaining the undepreciated portion of its capital, net of the adjustment cost, or the choice of continuing and investing the optimal amount of capital.

5. Keep on iterating until  $\sup \left| \frac{V_{t+1}(w, z, k, s) - V_t(w, z, k, s)}{V_t(w, z, k, s)} \right| < 10.0^{-6}$ . Denote the latest value function as  $V_\infty(w, z, k, s)$ .

## A.2 Entry

1. Define a grid for the signal. Denote it as  $\Psi_q$ . Let also  $W(s_n | q)$  indicate the probability that the first draw of the idiosyncratic shock is  $s_n$ , conditional on the signal being  $q$ .

2. For all triplets  $(w, z, q)$  on the grid, compute the value of entering as

$$V_e(w, z, q) = \max_{k' \in \Psi_k} -k' + \frac{1}{R} \sum_j \sum_i \sum_n V_\infty(w_i, z_j, k', s_n) W(s_n | q) J(w_i | w, z, z_j) G(z_j | z).$$

3. For all grid points  $z$ , construct a bi-dimensional cubic spline interpolation of  $V_e(w, z, q)$  over the dimensions  $(w, q)$ . For all pairs  $\tilde{w}, \tilde{q}$ , denote the value of entering as  $\tilde{V}_e(\tilde{w}, z, \tilde{q})$ .
4. Define  $\tilde{q}_e(\tilde{w}, z)$  as the value of the signal which makes prospective entrants indifferent between entering and not. That is,  $\tilde{V}_e(\tilde{w}, z, \tilde{q}_e(\tilde{w}, z)) = c_e$ .

### A.3 Simulation

- Given the current firm distribution  $\Gamma_t$  and aggregate shock  $z_t$ , compute the equilibrium wage  $\tilde{w}_t$  by equating the labor supply equation  $L_s(w) = w^\gamma$  to the labor demand equation

$$L_t^d(w) = \left( \frac{z_t \theta (1-\alpha)}{w} \right) \sum_m \sum_n \left[ s_n k_m^{\alpha\theta} \right]^{\frac{1}{1-\theta(1-\alpha)}} \Gamma_t(s_n, k_m).$$

- For all  $z' \in \Psi_z$ , compute the conditional wage forecast  $\tilde{w}_{t+1}(z')$  as follows:

$$\log[\tilde{w}_{t+1}(z')] = \hat{\beta}_0 + \hat{\beta}_1 \log(\tilde{w}_t) + \hat{\beta}_2 \log(z') + \hat{\beta}_3 \log(z_t).$$

For every  $z'$ , there will be contiguous grid points  $(w_i, w_{i+1})$  such that  $w_i \leq \tilde{w}_{t+1}(z') \leq w_{i+1}$ . Now let  $J_{t+1}(w_i|z') = 1 - \frac{\tilde{w}_{t+1}(z') - w_i}{w_{i+1} - w_i}$ ,  $J_{t+1}(w_i|z') = \frac{\tilde{w}_{t+1}(z') - w_i}{w_{i+1} - w_i}$ , and  $J_{t+1}(w_j|z') = 0$  for all  $j$  such that  $j \neq i$  and  $j \neq i+1$ ;

- For all pairs  $(k, s)$  on the grid such that  $\Gamma_t(k, s) > 0$ , the optimal choice of capital  $k'(\tilde{w}_t, z_t, k, s)$  is the solution to the following problem:

$$\begin{aligned} \max_{k' \in \Psi_k} & \pi(\tilde{w}_t, z_t, k, s) - x - c_0 k \chi - c_1 \left( \frac{x}{k} \right)^2 k \\ & + \frac{1}{R} \sum_j \sum_i \sum_n V_\infty(w_i, z_j, k', s_n) H(s_n|s) J_t(w_i|z_j) G(z_j|z_t), \\ \text{s.t. } & x = k' - k(1-\delta), \\ & \pi(\tilde{w}_t, z_t, k, s) = \frac{1 - (1-\alpha)\theta}{(1-\alpha)\theta} \tilde{w}_t^{-\frac{\theta(1-\alpha)}{1-\theta(1-\alpha)}} [(1-\alpha)\theta s z_t k^{\alpha\theta}]^{\frac{1}{1-\theta(1-\alpha)}}, \\ & \chi = 1 \text{ if } k' \neq k \text{ and } \chi = 0 \text{ otherwise .} \end{aligned}$$

- There will be contiguous elements of the signal grid  $(q^*, q^{**})$  such that  $q^* \leq \tilde{q}_e(\tilde{w}_t, z_t) \leq q^{**}$ .

- For all  $q \geq q^{**}$ , the initial capital of entrants  $k'_e(\tilde{w}_t, z_t, q)$  solves the following problem:

$$\max_{k'_e \in \Psi_k} -k_m + \frac{1}{R} \sum_j \sum_i \sum_n V_\infty(w_i, z_j, k'_e, s_n) W(s_n|q) J_t(w_i|z_j) G(z_j|z_t).$$

- We can easily compute the distribution of the idiosyncratic shock conditional on  $\tilde{q}_e \equiv \tilde{q}_e(\tilde{w}_t, z_t)$ , denoted as  $\tilde{W}(s_n|\tilde{q}_e)$  and then compute the optimal capital  $k'_e(\tilde{w}_t, z_t, \tilde{q}_e)$  as the solution to:

$$\max_{k'_e \in \Psi_k} -k_m + \frac{1}{R} \sum_j \sum_i \sum_n V_\infty(w_i, z_j, k'_e, s_n) \tilde{W}(s_n|\tilde{q}_e) J_t(w_i|z_j) G(z_j|z_t).$$

5. Draw the aggregate productivity shock  $z_{t+1}$ ;
6. Determine the distribution at time  $t+1$ . For all  $(k, s)$  such that  $V_\infty(\tilde{w}_{t+1}(z_{t+1}), z_{t+1}, k, s) = 0$ , then  $\Gamma_{t+1} = 0$ . For all other pairs,

$$\Gamma_{t+1}(k, s) = \sum_m \sum_n \Gamma_t(k_m, s_n) H(s|s_n) \Upsilon_{m,n}(w_t, z_t, k) + \mathcal{E}_{t+1}(k, s),$$

where

$$\Upsilon_{m,n}(w_t, z_t, k) = \begin{cases} 1 & \text{if } k'(w_t, z_t, k_m, s_n) = k \\ 0 & \text{otherwise.} \end{cases}$$

and  $\mathcal{E}_{t+1}(k, s) = M \sum_{i: q_i \geq q^{**}} H(s|q_i) Q(q_i) \Xi_{m,i}(w_t, z_t, k) + M \tilde{H}(s|q^e) \tilde{Q}(q^e) \tilde{\Xi}_{m,i}(w_t, z_t, k)$   
where

$$\Xi_{m,i}(w_t, z_t) = \begin{cases} 1 & \text{if } k'(w_t, z_t, q_i) = k \\ 0 & \text{otherwise.} \end{cases}$$

#### A.4 More Accuracy Tests of the Forecasting Rule

Forecasting errors are essentially unbiased – the mean error is -0.00017% of the forecasting price – and uncorrelated with the price (the correlation coefficient between the two series is -0.0456) and the aggregate shock (-0.0045).



Figure 18: Accuracy of the Forecasting Rule.

In the left panel of Figure 18 is the scatter plot of the forecasting errors against the market clearing price. In the right panel is the time series of the forecasting error. The good news is that errors do not cumulate.

## References

- ABRAHAM, A., AND K. WHITE (2006): “The Dynamics of Plant–Level Productivity in US Manufacturing,” *Census Bureau Working paper #06-20*.
- BACHMAN, R., AND C. BAYER (2013): “Wait-and-see Business Cycles?,” *Journal of Monetary Economics*, 60(6), 704–719.
- BASU, S., AND J. FERNALD (1997): “Returns to Scale in US Production: Estimates and Implications,” *Journal of Political Economy*, 105(2), 249–283.
- BEAUDRY, P., AND A. GUAY (1996): “What Do Interest Rates Reveal About the Functioning of Real Business Cycle Models?,” *Journal of Economic Dynamics and Control*, 20, 1661–1682.
- BILBIE, F., F. GHIRONI, AND M. MELITZ (2012): “Endogenous Entry, Product Variety, and Business Cycles,” *Journal of Political Economy*, 120(2), 304–345.
- BLOOM, N., M. FLOETOTTO, N. JAIMOVICH, I. SAPORTA-EKSTEN, AND S. TERRY (2014): “Really Uncertain Business Cycles,” NBER Working Paper # 18245.
- CABRAL, L., AND J. MATA (2003): “On the Evolution of the Firm Size Distribution: Facts and Theory,” *American Economic Review*, 93(4), 1075–1090.
- CAMPBELL, J. (1998): “Entry, Exit, Embodied Technology, and Business Cycles,” *Review of Economic Dynamics*, 1, 371–408.
- CAO, S. (2007): “Three Essays on Capital Adjustment, Reallocation and Aggregate Productivity,” PhD Dissertation, University of Texas at Austin.
- CASTRO, R., G. L. CLEMENTI, AND Y. LEE (2015): “Cross-Sectoral Variation in The Volatility of Plant-Level Idiosyncratic Shocks,” *Journal of Industrial Economics*, 63(1), 1–29.
- CETORELLI, N. (2014): “Surviving Credit Market Competition,” *Economic Inquiry*, 52(1), 320–340.
- CHATTERJEE, S., AND R. COOPER (2014): “Entry and Exit, Product Variety and the Business Cycle,” *Economic Inquiry*, 52(4), 1466–1484.
- CHOI, H. (2014): “The Role of Establishment Heterogeneity in the Recovery from Sudden Stops,” Working Paper, Monash University.

- COAD, A. (2009): *The Growth of Firms. A Survey of Theories and Empirical Evidence.* Edward Elgar Publishing.
- COGLEY, T., AND J. NASON (1995): “Output Dynamics in Real Business Cycle Models,” *American Economic Review*, 85(3), 492–511.
- COOLEY, T., AND E. PRESCOTT (1995): “Economic Growth and Business Cycles,” in *Frontiers of Business Cycle Research*, ed. by T. Cooley, pp. 1–38. Princeton University Press.
- COOLEY, T., AND V. QUADRINI (2001): “Financial Markets and Firm Dynamics,” *American Economic Review*, 91(5), 1286–1310.
- COOPER, R., AND J. HALTIWANGER (2006): “On the Nature of Capital Adjustment Costs,” *Review of Economic Studies*, 73(3), 611–633.
- COOPER, R., AND J. WILLIS (2014): “Discounting: Investment Sensitivity and Aggregate Implications,” Working Paper, Penn State.
- DEN HAAN, W. J. (2010): “Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents,” *Journal of Economic Dynamics and Control*, 34, 79–99.
- D’ERASMO, P. (2011): “Investment and Firm Dynamics,” Working Paper, University of Maryland.
- DEVEREUX, M., A. HEAD, AND B. LAPHAM (1996): “Aggregate Fluctuations with Increasing Returns to Specialization and Scale,” *Journal of Economic Dynamics and Control*, 20, 627–656.
- DUNNE, T., M. ROBERTS, AND L. SAMUELSON (1988): “Patterns of Firm Entry and Exit in US Manufacturing Industries,” *Rand Journal of Economics*, 19, 495–515.
- (1989): “The Growth and Failure of US Manufacturing Plants,” *Quarterly Journal of Economics*, 94, 671–698.
- EISFELDT, A., AND A. RAMPINI (2006): “Capital Reallocation and Liquidity,” *Journal of Monetary Economics*, 53, 369–399.
- EVANS, D. (1987): “Tests of Alternative Theories of Firm Growth,” *Journal of Political Economy*, 95(4), 657–674.

- FIORITO, R., AND G. ZANELLA (2012): “The Anatomy of the Aggregate Labor Supply Elasticity,” *Review of Economic Dynamics*, 15, 171–187.
- FOSTER, L., J. HALTIWANGER, AND C. KRIZAN (2001): “Aggregate Productivity Growth: Lessons from Microeconomic Evidence,” in *New Developments in Productivity Analysis*, ed. by C. Hulten, E. Dean, and M. Harper, pp. 303–363. University of Chicago Press.
- FOSTER, L., J. HALTIWANGER, AND C. SYVVERSON (2008): “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity of Profitability?,” *American Economic Review*, 98(1), 394–425.
- (2015): “The Slow Growth of New Plants: Learning about Demand?,” *Economica*, Forthcoming.
- HALL, B. (1987): “The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector,” *Journal of Industrial Economics*, 35(4), 583–606.
- HALTIWANGER, J. (1997): “Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence,” *Federal Reserve Bank of St. Louis Review*, 79, 55–77.
- (2011): “Firm Dynamics and Productivity Growth,” *European Investment Bank Papers*, 16(1), 116–136.
- HOPENHAYN, H. (1992): “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 60(5), 1127–1150.
- JAIMOVIC, N., AND M. FLOETOTTO (2008): “Firm Dynamics, Markup Variations, and The Business Cycle,” *Journal of Monetary Economics*, 55, 1238–1252.
- JARMIN, R., AND J. MIRANDA (2002): “The Longitudinal Business Database,” CES Working Paper # 02-17, US Census Bureau.
- KEHRIG, M. (2015): “The Cyclical Nature of the Productivity Distribution,” Working Paper, University of Texas at Austin.
- KHAN, A., AND J. K. THOMAS (2008): “Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics,” *Econometrica*, 76(2), 395–436.
- KING, R., C. PLOSSER, AND S. REBELO (1988): “Production, Growth and Business Cycles I: The Basic Neoclassical Model,” *Journal of Monetary Economics*, 21, 195–232.

- KRUSELL, P., AND A. SMITH (1998): “Income and wealth heterogeneity in the macroeconomy,” *Journal of Political Economy*, 106, 867–896.
- LEE, Y. (2005): “The Importance of Reallocations in Cyclical Productivity and Returns to Scale: Evidence from Plant-level Data,” Working paper # 05–09, Federal Reserve Bank of Cleveland.
- LEE, Y., AND T. MUKOYAMA (2013): “Entry, Exit, and Plant-level Dynamics over the Business Cycle,” Working paper, University of Virginia.
- (2015): “Entry and Exit of Manufacturing Plants over the Business Cycle,” *European Economic Review*, 77, 20–27.
- MCGRATTAN, E. (1999): “Application of weighted residual methods to dynamic economic models,” in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott. Oxford University Press.
- MCGRATTAN, E., AND E. PRESCOTT (2010): “Unmeasured Investment and the Puzzling US Boom in the 1990s,” *American Economic Journal: Macroeconomics*, 2(4), 88–123.
- ROGERSON, R., AND J. WALLENIUS (2009): “Micro and Macro Elasticities in a Life Cycle Model with Taxes,” *Journal of Economic Theory*, 144, 2277–2292.
- SAMANIEGO, R. (2008): “Entry, Exit and Business Cycle in a General Equilibrium Model,” *Review of Economic Dynamics*, 11, 529–541.
- SEDLACEK, P. (2015): “Lost Generations of Firms and Aggregate Labor Market Dynamics,” Working Paper, University of Bonn.
- TAUCHEN, G. (1986): “Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions,” *Economics Letters*, 20, 177–181.
- VERACIERTO, M. (2002): “Plant–Level Irreversible Investment and Equilibrium Business Cycles,” *American Economic Review*, 92(1), 181–197.
- WINBERRY, T. (2014): “Lumpy Investment, Business Cycles, and Stimulus Policy,” Working Paper, Princeton.