DISCUSSION PAPER SERIES

DP21391

THE EMPLOYMENT CONCENTRATION CHANNEL OF MONETARY POLICY

Guido Ascari, Andrea Colciago and Marco Membretti

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

THE EMPLOYMENT CONCENTRATION CHANNEL OF MONETARY POLICY

Guido Ascari, Andrea Colciago and Marco Membretti

Discussion Paper DP21391
Published 15 April 2026
Submitted 14 April 2026

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

THE EMPLOYMENT CONCENTRATION CHANNEL OF MONETARY POLICY

Abstract

Under monetary tightenings, employment at small, high-churn firms contracts more than at large incumbents, raising the employment share of large firms. A mixed-frequency BVAR on U.S. data (1982–2018) shows that tightenings reduce firm entry and new-entrant hiring, severing inflows into small firms, while higher exit destroys small-firm employment. Large incumbents are comparatively insulated, rarely exiting and exhibiting weak sensitivity to entry conditions. This mechanism raises employment concentration, defining an employment concentration channel of monetary policy. An estimated structural model with heterogeneous firms, endogenous entry and exit, and equilibrium unemployment matches this effect, showing that the concentration channel is quantitatively important in accounting for the empirical output-inflation trade-off.

JEL Classification: C13, E32

Keywords: Monetary policy

Guido Ascari - guido.ascari@unipv.it
University of Pavia, De Nederlandsche Bank and CEPR

Andrea Colciago - andreacolciago@gmail.com
Deutsche Bundesbank, University of Amsterdam, University of Milano-Bicocca

Marco Membretti - marco.membretti@gmail.com
European Commission - Joint Research Centre, Ispra

The Employment Concentration Channel of Monetary Policy*

Guido Ascari†
University of Pavia
De Nederlandsche Bank
CEPR

Andrea Colciago‡
Deutsche Bundesbank
University of Amsterdam
University of Milano-Bicocca

Marco Membretti§
European Commission, JRC, Ispra

April 15, 2026

Abstract

Under monetary tightenings, employment at small, high-churn firms contracts more than at large incumbents, raising the employment share of large firms. A mixed-frequency BVAR on U.S. data (1983–2018) shows that tightenings reduce firm entry and new-entrant hiring, severing inflows into small firms, while higher exit destroys small-firm employment. Large incumbents are comparatively insulated, rarely exiting and exhibiting weak sensitivity to entry conditions. This mechanism raises employment concentration, defining an employment concentration channel of monetary policy. An estimated structural model with heterogeneous firms, endogenous entry and exit, and equilibrium unemployment matches this effect, showing that the concentration channel is quantitatively important in accounting for the empirical output-inflation trade-off.

Keywords: Monetary policy; employment concentration; unemployment; heterogeneous firms; BVAR.

JEL classification: E52, E32, C13..

---

*Ascari and Membretti acknowledge generous support from the Ministry of the University, MIUR, PRIN 2020, Prot. 2020LKAZAH. The views expressed are those of the authors and do not necessarily reflect those of De Nederlandsche Bank, the Bundesbank, or the European Commission. Any remaining errors are ours.

†Corresponding author, Department of Economics and Management, University of Pavia, Via San Felice 5, Pavia 27100 Italy, and De Nederlandsche Bank, Frederiksplein 61, 1017 XL Amsterdam. Email: `guido.ascari@unipv.it.`

‡Department of Economics (DEMS), University of Milano Bicocca, Piazza dell'Ateneo Nuovo 1, Milano 20126, Italy. Email:`andreacolciago@gmail.com`

§European Commission - Joint Research Centre, Via E. Fermi, 2749, Ispra VA 21027, Italy. Email: `marco.membretti@gmail.com`

1 Introduction

Contractionary monetary policy does not merely shrink the economy: it reshapes it. When the central bank tightens, small and young firms, which operate closer to their exit threshold and depend disproportionately on new entrants to sustain their employment share, are hit on two fronts simultaneously: entry collapses, severing the flow of new jobs into the small-firm segment, while exit accelerates, destroying existing ones. Large established incumbents, by contrast, are insulated: they rarely exit under normal conditions, and their employment responds little to fluctuations in the entry margin. The result is a persistent increase in employment concentration, as the employment share of large incumbents rises. This increase in concentration outlasts the initial shock and persistently affects aggregate dynamics .

This paper makes three contributions. First, it documents this mechanism empirically. Second, it develops a structural model with heterogeneous firms, endogenous entry and exit, and equilibrium unemployment that rationalizes it. Third, it estimates the model and shows that (i) Matching the empirical degree of employment concentration and the observed size-dependent exit rates is essential for generating the increase in employment concentration following a monetary policy shock, and (ii) this concentration channel is quantitatively important for the empirical output–inflation trade-off faced by monetary policy.

Empirical evidence. We document the employment concentration channel using a mixed-frequency Bayesian VAR estimated on U.S. data from 1983 to 2018, combining standard quarterly macroeconomic variables with annual firm-level outcomes from the Business Dynamics Statistics. Following a contractionary monetary policy shock, employment concentration, measured as the share of workers at firms with 500 or more employees, rises persistently, peaking around twelve quarters after the shock before gradually reverting. The firm entry rate declines sharply and persistently. These aggregate responses mask substantial heterogeneity in job flows: job creation by new entrants falls sharply and persistently, while job creation by incumbent firms displays no significant decline. Job destruction rises almost entirely among small firms, through both elevated exit and contraction of continuing small businesses. Large firms are essentially unaffected on both margins. The combination of weaker entry and high small-firm exit leads to a persistent decline in net entry and a gradual rise in employment concentration among large incumbents.

We quantify the central bank's output-inflation trade-off using the Phillips multiplier of Barnichon and Mesters (2021), defined as the ratio of cumulative inflation changes to cumulative output changes following a policy shock.

Structural model. To rationalize these findings, we develop a structural model with heterogeneous firms, endogenous entry, endogenous exit rates, equilibrium unemployment, and nominal rigidities. Firms have heterogeneous permanent productivity and hire multiple workers under decreasing returns to labor. Small firms, which draw lower permanent productivity, operate closer to their exit threshold and exit at substantially higher rates than large incumbents. New entrants account for a disproportionate share of gross job creation relative to their employment share, consistent with Haltiwanger et al. (2013). The model is estimated using a limited-information minimum distance approach, targeting both the empirical impulse responses to a monetary policy shock and key moments of the stationary firm size distribution from the Business Dynamics Statistics.

This joint discipline ensures that both the stationary distribution of firms and the dynamic transmission of monetary policy are tightly anchored to the data.

The estimated model replicates the persistent rise in employment concentration, the asymmetric behavior of job flows across firm sizes, and the declining Phillips multiplier observed in the data. Consistent with the data, a contractionary shock reduces entry and raises small-firm exit, shifting the employment distribution toward large incumbents and generating a rise in employment concentration that outlasts the initial shock.

Counterfactual analysis. To isolate the structural features driving this mechanism, we consider two counterfactual economies. In the first, we shut down endogenous exit by imposing a size-independent exit rate, thereby attenuating the contraction in the mass of active firms following a tightening. In the second, we reduce steady-state employment concentration, approximating an economy in which production is more evenly distributed across firms. In both cases, the rise in employment concentration following a contractionary shock is substantially muted, because the differential contraction across firm cohorts is weakened.

When exit rates are equalized, small firms no longer bear a disproportionate share of the contraction, limiting the change in employment concentration. When large firms account for a negligible share of employment, the economy is close to a homogeneous-firm environment, so the distribution of employment across firms changes little in response to the shock. These counterfactuals disentangle two distinct forces. The first governs how much productive capacity contracts, through changes in the mass of active firms. The second governs how a given contraction translates into aggregate outcomes, by altering which firms account for production and therefore how strongly prices and quantities adjust. When employment is less concentrated and small firms account for a larger share of activity, exit and entry dynamics induce a stronger contraction in effective productive capacity, dampening the price response and shifting the adjustment toward quantities. Conversely, when employment is concentrated among large incumbents, the same shock affects aggregate capacity less, and a larger fraction of the adjustment occurs through prices. We estimate each restricted model using the same procedure used for the baseline model and find that both counterfactuals generate an output–inflation trade-off that is quantitatively and dynamically inconsistent with the data. This implies that the employment concentration channel is central to the model's ability to reproduce the empirical Phillips multiplier, both in magnitude and in its dynamic profile.

Policy implications. The output-inflation trade-off faced by monetary policy is not a primitive of the economy: it is endogenous to firm composition, because it depends on how production and employment are distributed across firms. In our framework, firm composition determines both how much productive capacity contracts following a monetary tightening and how this contraction translates into price and quantity adjustment. The counterfactual analysis shows that these differences are quantitatively important, highlighting that changes in firm composition can modify the trade-off faced by the central bank.

This conclusion has direct policy relevance. Between 1983 and 2018, the period we consider, we observed a a secular decline in business dynamism and firm entry, while the employment share of large and established firms has risen from 45.5 percent to 52.5 percent. Our framework implies that this structural transformation has not been neutral for monetary transmission: as the economy has shifted away from fragile small firms toward large incumbents, the employment concentration channel documented here has itself changed, and with it the output-inflation trade-off that the central bank faces. Monetary policy transmission is therefore inherently linked to market structure, and changes in firm composition can reshape the output–inflation trade-off faced by central banks.

2 Related Literature

This paper contributes to four strands of the macroeconomic literature.

Firm dynamics and labor market frictions. A growing literature studies how firm dynamics interact with labor market frictions to shape unemployment dynamics and aggregate fluctuations. Empirically, Haltiwanger et al. (2013) and Fort et al. (2013) document the strong job creation contribution of young firms over the long run and the pronounced cyclicality of employment at young and small businesses. On the theory side, Elsby and Michaels (2013) develop a structural model with endogenous, size-dependent job destruction, highlighting how firm-level dynamics affect unemployment flows. Sedláček (2020) shows that missing entrants generated persistent effects on unemployment and output following the great recession, while Siemer (2014) emphasizes that shocks suppressing firm entry can lead to prolonged downturns through reduced job creation. Bilal et al. (2022) study joint firm and worker dynamics with on-the-job search, highlighting the role of worker reallocation across firms. Colciago et al. (2025) show that a search and matching model with frictional firm entry matches the empirical cyclicality of profits and hours worked following aggregate supply shocks, unlike standard frameworks with frictionless entry. We contribute to this literature by developing and estimating a model with endogenous entry, size-dependent exit, and equilibrium unemployment to study the transmission of monetary policy shocks, showing that policy-induced distortions to entry and exit generate persistent changes in employment concentration.

Heterogeneous firm responses to monetary policy shocks. The asymmetric effects of monetary policy across firms of different sizes have attracted increasing attention since the seminal contribution of Gertler and Gilchrist (1994), which documents a larger elasticity of small firms' sales to monetary policy tightening. Using UK micro data, Bahaj et al.

(2022) find stronger employment responses at younger and more leveraged firms. Popov and Steininger (2023) show that monetary policy shocks affect market competition in the euro area by disproportionately influencing sales growth at small firms. Singh et al. (2022) document asymmetric responses of the firm size distribution to monetary policy shocks within a heterogeneous firm framework, but abstract from the extensive margin of adjustment, namely entry and exit. We contribute to this literature by providing a unified empirical and structural analysis of employment allocation across firms of different sizes, quantifying the role of entry and size-dependent exit in shaping firm composition and aggregate dynamics following monetary policy shocks.

Firm entry, exit, and aggregate shocks. A growing literature, building on Melitz (2003), Ghironi and Melitz (2005), Bilbiie et al. (2012), Jaimovich and Floetotto (2008), Colciago and Etro (2010), Clementi and Palazzo (2016), Hamano and Zanetti (2017), and Rossi (2019), studies how firm entry, exit, and product variety shape business cycle dynamics. Early contributions, such as Bilbiie et al. (2007) and Bergin and Corsetti (2008), focused on monetary transmission in the presence of an extensive margin of investment. Lewis and Poilly (2012) and Bilbiie (2020) revisit these mechanisms in settings with homogeneous firms, while Colciago and Silvestrini (2022) extend the analysis to environments with firm heterogeneity. We contribute to this literature by embedding both entry and size-dependent exit into a heterogeneous-firm model with equilibrium unemployment and nominal rigidities, estimated to match the empirical responses of firm and labor market aggregates to monetary policy shocks, and showing how firm composition shapes the output-inflation trade-off.

The output-inflation trade-off and the Phillips multiplier. Barnichon and Mesters (2021) introduce the Phillips multiplier as a non-parametric statistic to characterize the central bank's output-inflation trade-off, showing that it declines persistently over the horizon: monetary tightening achieves temporary disinflation only at the cost of a very persistent output loss, and that this declining pattern reflects primarily the anchoring of inflation expectations rather than a flattening of the Phillips curve slope. Debortoli et al. (2023) document asymmetries in the trade-off across the business cycle, while Gabriel (2023) reports asymmetries over time. Our paper contributes to this literature by providing a new structural mechanism that accounts for the empirically observed Phillips multiplier. We show that the output-inflation trade-off is endogenous to firm composition: the degree of employment concentration among large incumbents and the size-dependence of exit jointly determine the multiplier, and abstracting from either feature, even within our estimated model, generates multipliers that are quantitatively inconsistent with the data. To our knowledge, this is the first paper to provide a structural, supply-side explanation for the Phillips multiplier rooted in the firm size distribution and the entry-exit margin of monetary transmission.

3 Empirical Analysis

3.1 Data, Methodology, and Identification

This section presents empirical evidence on the macroeconomic effects of monetary policy shocks on aggregate activity, firm entry, and employment concentration. We document how monetary tightening differentially affects fragile small firms and large established incumbents, generating a persistent shift in the employment shares of the two groups.

We estimate a mixed-frequency Bayesian VAR (BVAR) that combines standard financial and macroeconomic variables with measures of business dynamism and employment distribution, allowing us to capture both aggregate and distributional responses.[1] Since the key measures of business dynamism and employment distribution from the Business Dynamics Statistics (BDS) are available only at the annual frequency, the mixed-frequency framework allows us to incorporate these variables within a standard quarterly macroeconomic setting.

In the mixed-frequency BVAR, the quarterly block includes: real GDP, the unemployment rate, the year-on-year change in the GDP deflator, the market yield on 1-year U.S. Treasury securities (GS1), and the GZ excess bond premium (EBP). All quarterly variables are sourced from the Federal Reserve Economic Data (FRED). The annual block contains the employment share of firms with at least 500 employees, which captures employment concentration, and the entry rate of new firms, defined as the ratio of newborn firms to the total number of firms (incumbents and entrants). Both variables are sourced from the Business Dynamics Statistics (BDS) database over the period 1983–2018.[2] BDS provides annual data

[1] We employ a Minnesota prior for the VAR parameters and implement the estimation using the MATLAB toolkit developed by Canova and Ferroni (2020).

[2] The annual BDS data correspond to March (Q1) of each year. To map these annual observations into a quarterly frequency within the state-space representation, the Kalman filter requires a window of three

on establishment-level employment and job flows, disaggregated by parent-firm size. A firm is defined as the collection of all its establishments. We reorganize firms in BDS into three groups. The first group consists of new entrants, which by construction have zero initial employment. The second group comprises small firms with fewer than 500 employees, while the third group consists of large firms with 500 or more employees.[3]

Variable	**Sign**	**Quarters**
1-year rate	+	4
GDP	-	4
Inflation	-	4
M1	-	4
EBP	+	4

Table 1: Sign restrictions for a contractionary monetary policy shock, valid for four quarters.

Monetary policy shocks are identified using sign restrictions. Table 1 summarizes the imposed restrictions and their duration. We follow standard assumptions that a monetary tightening raises the 1-year Treasury yield and the excess bond premium, and reduces GDP, the price index, and M1.[4]

3.2 Empirical Findings

3.2.1 Firm composition and the Labor Market

Figure 1 displays the impulse responses of key variables to a contractionary monetary policy shock. Solid lines show median responses, while shaded areas denote 68 percent credible intervals.

Consistent with the imposed sign restrictions, output and inflation decline over the first four quarters and continue to do so beyond the restricted horizon, indicating persistent responses. While inflation reaches its trough after a few quarters and then gradually reverts

preceding quarterly observations. Consequently, while the reported sample period is 1983–2018, the quarterly variables are initialized in 1982-Q2 to facilitate the frequency conversion of the first annual data point.

[3]Appendix A provides detailed definitions of all variables and describes their treatment.

[4]As a robustness check, Appendix B presents an alternative identification based on a monthly Proxy SVAR using high-frequency external instruments and shows that our main findings are robust across identification approaches.

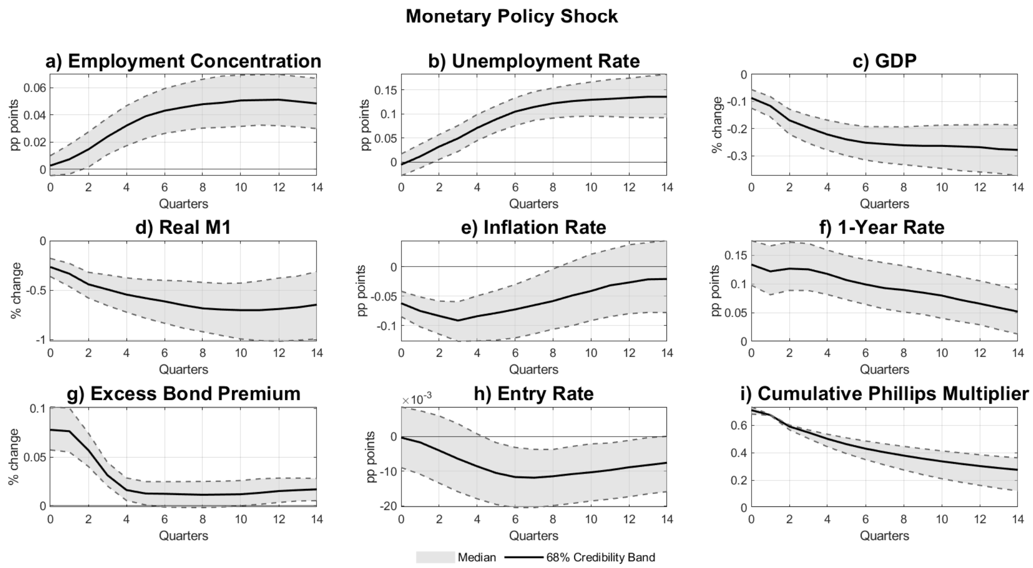

Figure 1: Baseline VAR: IRFs to a monetary policy shock (68% credible bands in gray).

toward its pre-shock level, output remains persistently below trend. The unemployment rate rises persistently in response to the shock, as well as employment concentration—measured as the share of workers employed at firms with at least 500 employees— which peaks around twelve quarters after the shock before gradually returning toward its initial level. The firm entry rate also declines persistently over several quarters.

These results indicate that a temporary monetary tightening can have lasting effects on labor market dynamics and the firm size distribution. In particular, the distribution of employment shifts away from small and fragile firms toward larger established firms following the shock. We refer to this mechanism as the *employment concentration channel of monetary policy*. The following sections examine the trade-off between inflation and output and investigate the mechanisms underlying the employment concentration pattern by analyzing job flows.

3.2.2 The Trade-Off Between Inflation and Output

We quantify the central bank's trade-off between output and inflation using a modified version of the Phillips multiplier, as introduced by Barnichon and Mesters (2021). This statistic captures how effectively the central bank can convert changes in the policy rate into changes in output or inflation. Specifically, the average trade-off is defined as the ratio of the average change in inflation to the average change in output resulting from a one standard deviation shock to the policy rate. Formally, it is defined as:

$$T_h = \frac{\frac{\partial \tilde{\pi}_{0:h}}{\partial \epsilon^R}}{\frac{\partial \tilde{y}_{0:h}}{\partial \epsilon^R}}, \quad h \geq 0, \tag{1}$$

where the tilde variables denote the average value of some variable over horizon h:

$$\tilde{x}_{t:t+h} = \frac{1}{h+1} \sum_{j=0}^{h} x_{t+j}.$$

We compute T_h using the impulse responses of inflation and output to monetary policy shocks obtained from our BVAR model. The median values of T_h for the different horizons h are reported in the bottom right panel of Figure 1.

On impact, T_h is approximately 0.75, and then it decreases persistently. This pattern implies that the impact response of inflation to a monetary policy contraction is lower than the one of output, and the ratio between the two cumulative responses keeps decreasing as the horizon h grows, as a result of the different persistence of the impulse response functions of output and inflation in Figure 1. In other words, the central bank can induce a temporary reduction of inflation at the cost of a very persistent cost in terms of output, and the trade-off worsens over the horizon h.

3.2.3 Job Flows

To shed light on the sources of the rise in unemployment and employment concentration following a monetary policy shock, we use our estimated monetary policy shocks from the BVAR as input in local projections à la Jordà (2005) to estimate the impulse responses for aggregate job flows, as well as for job flows by firm category.

Following Colciago et al. (2019), we decompose the job creation rate (JCR) into an extensive and an intensive margin. The extensive margin captures job creation due to firm entry, while the intensive margin captures job creation due to the expansion of incumbent firms. In any period, aggregate job creation can be written as:

$$JC_t = JC_t^{new} + JC_t^{exp}, \tag{2}$$

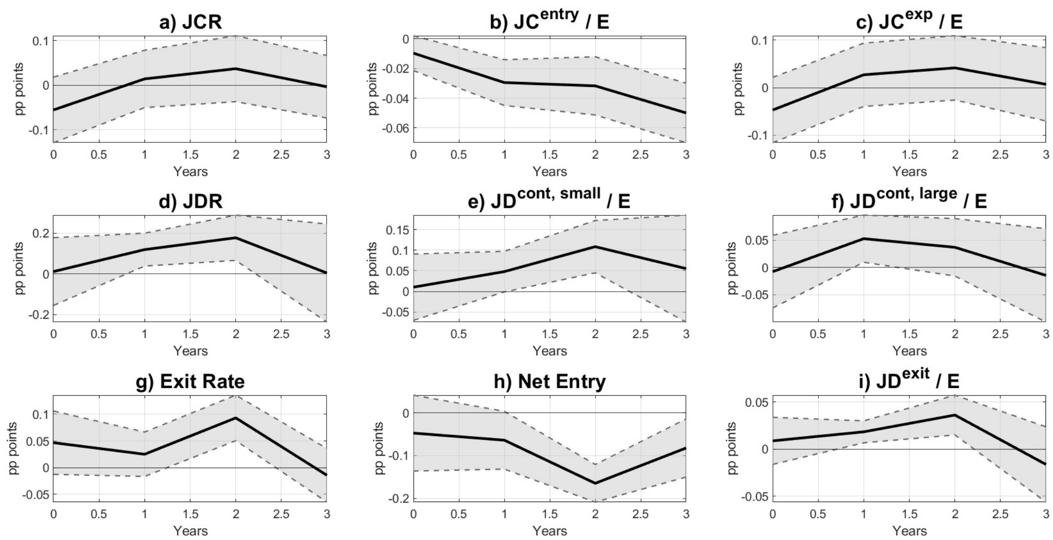

Figure 2: IRFs to a MP shock (68% confidence interval).

where JC_t^{new} denotes jobs created by new firms, and JC_t^{exp} those created by expanding incumbent firms. Dividing by total employment E_t, the aggregate job creation rate can be decomposed as:

$$JCR_t = \underbrace{\frac{JC_t^{new}}{E_t}}_{\text{extensive margin}} + \underbrace{\frac{JC_t^{exp}}{E_t}}_{\text{intensive margin}} . \tag{3}$$

Panels a)–c) of Figure 2 display this decomposition. Following a contractionary monetary policy shock, aggregate job creation (panel a) declines modestly and not significantly. This masks substantial heterogeneity: job creation by new entrants falls sharply and persistently (panel b), while job creation by incumbent firms initially falls and then rises, but not significantly so. Note that the response of incumbents shapes the overall dynamics of total job creation, given their relative size. This distinction is important because new entrants contribute disproportionately to gross job creation relative to their employment share, as emphasized by Haltiwanger et al. (2013).

We next turn to the job destruction margin and propose a finer decomposition by firm size. Jobs can be destroyed either through firm exit or through contraction by continuing firms. Let JD_t^{exit} denote jobs destroyed due to exit, and JD_t^{cont} those destroyed by contraction.

The aggregate job destruction rate can be written as:

$$JDR_t = \underbrace{\frac{JD_t^{exit}}{E_t}}_{\text{extensive margin}} + \underbrace{\frac{JD_t^{cont}}{E_t}}_{\text{intensive margin}} . \tag{4}$$

Job destruction by contracting firms can be split by firm size:

$$JD_t^{cont} = JD_t^{cont,S} + JD_t^{cont,L}, \tag{5}$$

where $JD_t^{cont,S}$ and $JD_t^{cont,L}$ denote job destruction at small and large firms, respectively. Hence, the aggregate job destruction rate can be expressed as:

$$JDR_t = \underbrace{\frac{JD_t^{exit}}{E_t}}_{\text{extensive margin}} + \underbrace{\frac{JD_t^{cont,S}}{E_t}}_{\text{intensive margin, small firms}} + \underbrace{\frac{JD_t^{cont,L}}{E_t}}_{\text{intensive margin, large firms}} . \tag{6}$$

Aggregate job destruction rises a few quarters after the shock (panel d) of Figure 2, driven by small firms, whose job destruction increases sharply after one year (panel e). In contrast, job destruction at large firms remains essentially unchanged (panel f). The rise in the exit rate (panel g) and job destruction due to exit (panel i) also occur almost entirely among small firms.[5] The combination of rising exit and weaker entry leads to a persistent decline in net entry (panel h).

These compositional effects are central to understanding the increase in employment concentration following a monetary policy shock. Small firms rely disproportionately on entry to generate the jobs needed to sustain their employment share. When entry declines and exit rises, job losses among small firms are not offset by new firm creation. Large firms, by contrast, are largely insulated from these margins. Over time, this asymmetry leads to a persistent increase in the concentration of employment in large firms.

4 The Model

To rationalize the findings from the previous sections, we propose a macroeconomic model with equilibrium unemployment and an endogenous firm size distribution. The model in-

[5]Job destruction due to exit of large firms, as well as their exit rate, display no meaningful response to the shock. For this reason, we do not report a decomposition along the size margin for these variables.

corporates real and nominal frictions in the labor and goods markets. The labor market features search frictions, following Mortensen and Pissarides (1994), where firms and workers engage in a matching process. On the production side, the structure includes three interconnected layers. At the bottom level, production firms have heterogeneous productivity and hire multiple workers to produce a homogeneous good. These firms face endogenous entry and exit dynamics, as in Hopenhayn (1992) and Clementi and Palazzo (2016), making the mass of producers endogenous. Establishing a new production firm requires an entry cost denominated in terms of the final good. Production firms supply their homogeneous goods to retailers, who sell differentiated goods to a final good producer. Retailers operate under nominal price rigidities à la Rotemberg (1982). At the top level, the final goods producer combines these differentiated goods into a composite good using a CES (Constant Elasticity of Substitution) production function, and sells it to consumers. On the demand side, a representative household works and owns shares in production firms. In what follows, we adopt the convention of omitting time subscripts for current-period variables, denoting lagged values with the subscript (-1) and forward values with a prime. When necessary, variables dated at time t are explicitly indexed by t.

4.1 Labor Market

Firms post vacancies to create new jobs. Matches between unemployed workers and vacancies are formed according to a constant returns to scale matching function:

$$M = \mu U^{1-\gamma} V^{\gamma}, \tag{7}$$

where U and V denote, respectively, the mass of unemployed workers and total vacancies before the creation of new matches. By defining $\theta = \frac{V}{U}$ as the tightness in the labor market, the matching function can be rewritten as: $M = \mu\theta^{\gamma-1}V$. The probability that a vacancy is filled is $q(\theta) = \frac{M}{V} = \mu\theta^{\gamma-1}$, while the probability of finding a job for an unemployed worker is $\phi(\theta) = \frac{M}{U} = \mu\theta^{\gamma} = q(\theta)\theta$. Firms and workers take both probabilities as given. As unemployed workers are matched with firms and their wages are set, they become productive.

4.2 Production

Next, we present the production side of the model, following the three-layer structure described above.

4.2.1 Production Firms

At time t, a positive mass of price-taking firms produces a homogeneous good with the following decreasing returns to labor production function:

$$y(Z, z, n) = Zzn^{\alpha}, \tag{8}$$

with $\alpha < 1$. The idiosyncratic productivity of a firm has two components: a permanent and a transitory one. Upon entry, a firm pays a fixed amount of output to draw its permanent idiosyncratic productivity component, Z, from a time-invariant distribution G. Productivity Z remains constant through the life cycle of the firm. The transitory component of the idiosyncratic productivity, is denoted by z. Its logarithm follows an autoregressive process of order 1, with parameter ρ_z, formally:

$$\log z' = \rho_z \log z + \epsilon^z, \tag{9}$$

where ϵ^z is an idiosyncratic productivity shock drawn from a normal distribution with mean zero and variance σ_z^2. This distinction is empirically relevant, as Sterk et al. (2021) show that ex ante heterogeneity across firms, rather than persistent ex post shocks, accounts for most of the differences in firm performance over the life cycle. Variable n_t denotes the mass of workers employed by an individual firm. Let $H(z'|z)$ denote the conditional distribution of z, and $\tilde{H}(z)$ its unconditional distribution. The distribution of operating firms across the three dimensions of heterogeneity is represented by the measure Λ on (Z, n, z), with total mass $N = \int d\Lambda$ equal to the mass of operating firms. Sums over firms are written as integrals with respect to Λ (for example, total employment at operating firms is $\int n \, d\Lambda$). Finally, let λ denote the vector of aggregate state variables, with $\Upsilon(\lambda'|\lambda)$ representing its law of motion. Following Clementi and Palazzo (2016), firms draw a fixed cost of production, c_f, from a time-invariant, log-normal distribution G with parameters μ_o and σ_o. A firm with permanent productivity Z exits the market when its value, defined as the discounted sum of future expected profits, turns negative. After observing their idiosyncratic productivity draw, firms choose their optimal employment adjustment policy to maximize their value function $F(Z, z, n)$. Separations occur at zero cost, while vacancy creation incurs a cost κ per vacancy. Hiring and wage bargaining take place simultaneously, with new hires becoming immediately productive. The mass of firms at any given time, N, is determined endogenously through entry and exit. Figure 3 illustrates the sequence of events within each period.

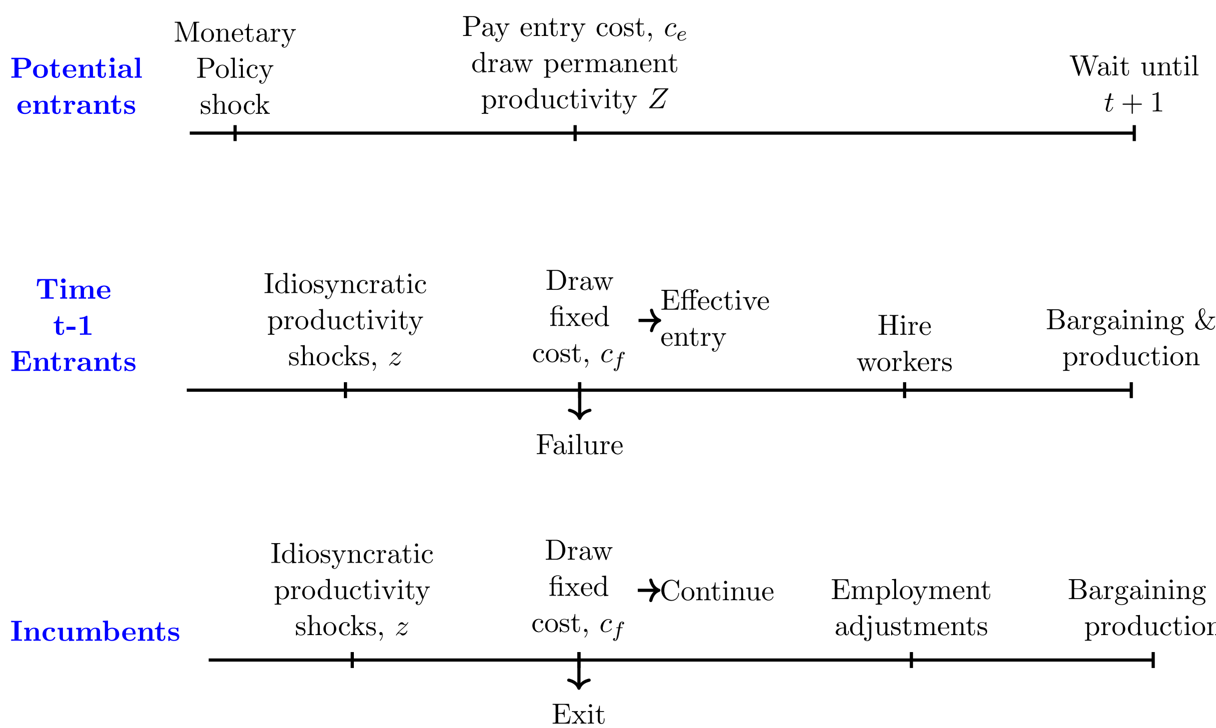

Figure 3: Timing.

Incumbents. We write the firm's optimization problem recursively. Consider a firm with permanent productivity Z, which at the beginning of a generic period-t receives an idiosyncratic productivity z. Let $F(Z, z, n)$ denote the beginning of period value of an incumbent firm, $w(Z, z, n)$ its wage, $v(Z, z, n)$ the mass of vacancies it posts, while $f(Z, z, n)$ the mass of workers that separate from the firm at time t. The variable $\tilde{\boldsymbol{\beta}}$ represents the real stochastic discount factor of the representative household that owns firms between period t and $t+1$, defined in section 4.3. To simplify notation, we drop the dependency of firm-specific variables on the triplet (Z, z, n) whenever it does not jeopardize clarity. The variable

ρ denotes the price of an individual firm's good relative to the price of the final good. The wage results from individual bargaining, described in Section 4.4. In response to a negative realization of the idiosyncratic productivity shock, an incumbent firm has three options: reduce its employment, freeze it until the next period, or exit from the market.

The value maximization problem of an incumbent firm reads as:

$$F(Z, z, n_{-1}) = \max_{n} \Bigg\{ \rho y - n\, w(Z, z, n) - \kappa \frac{(n - n_{-1})}{q} \mathbf{1}^{+} \\ + \int_{\mathcal{Z}} \int_{\mathcal{S}} \Big[\tilde{F}(Z, z', n) - c^f \Big] \, dG' dH'(z'|z) \Bigg\}, \tag{10}$$

where $\mathcal{S} = \{c^f : \tilde{F}(Z, z', n) > c^f\}$ represents the survival region for the following period - where the firm's value exceeds the fixed cost - $\mathcal{Z}$ denotes the full support of the idiosyncratic productivity shocks, and $\mathbf{1}^{+}$ is an indicator function equal to one if $n > n_{-1}$ and zero otherwise. The continuation value is defined as:

$$\tilde{F}(Z, z', n) = \tilde{\boldsymbol{\beta}} E_{\lambda'|\lambda}[F(Z', z', n)],$$

where the expectation is taken over the future aggregate state Z'. The maximization is subject to the evolution of firm-level employment:

$$n = \begin{cases} n_{-1} + qv & \text{if } n > n_{-1}, \\ n_{-1} - f & \text{if } n \leq n_{-1}. \end{cases}$$

Differentiating equation (10) with respect to n delivers the job creation condition (JCC):

$$J(Z, z, n) = \frac{\kappa}{q(Z, z, n)} \mathbf{1}^{+}, \tag{11}$$

where $J(Z, z, n)$ describes the marginal value of an additional worker to the firm:

$$J(Z, z, n) = \rho \alpha Z z n^{\alpha - 1} - \frac{\partial w(Z, z, n)}{\partial n} n - w(Z, z, n) + D(Z, z, n).$$

Following Elsby and Michaels (2013), we define $D(Z, z, n)$ as the marginal effect of current employment choices on the expected future value of the firm:

$$D(Z, z, n) \equiv \frac{\partial \left(\int_{\mathcal{Z}} \int_{\mathcal{S}} [\tilde{F}(Z, z', n) - c^f] dG' dH'(z'|z) \right)}{\partial n}. \tag{12}$$

Formally, this can be expressed as:

$$D(Z,z,n) = E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}} \int_{\mathcal{S}} \Big(\int_{\mathcal{Z}^{na}} J' dH'(z'|z) + \int_{\mathcal{Z}^{exp}} \frac{\kappa}{q'} dH'(z'|z) \Big) dG' \tag{13}$$

where $\mathcal{Z}^{na} = \{z : n = n_{-1}\}$ denotes the non-adjustment region, and $\mathcal{Z}^{exp} = \{z : n > n_{-1}\}$ represents the employment expansion region for the individual firm in the subsequent period (see Appendix C). The job creation condition (JCC) has a standard interpretation. The optimal employment level of a firm is such that the marginal value of labor, $J(Z,z,n)$, equals the cost of hiring an additional worker, given by the right-hand side of equation (11). The term $\frac{\partial w(Z,z,n)}{\partial n} n$ in the definition of the marginal value of a worker implies that the firm correctly anticipates that the bargained wage depends on its employment level.

Potential entrants. Potential entrants must pay a sunk entry cost c_e, measured in units of the final good, to draw the permanent component of their individual productivity. Firms enter the market up to the point where the sunk cost of entry is equal to the expected value of discounted future profits. Since the permanent component of idiosyncratic productivity is unknown ex-ante, entrants must compute the expected value of their discounted future profits over the distribution of permanent productivity components, G, and the unconditional distribution of idiosyncratic productivity shocks, $\tilde{H}(z)$. Formally, the expected value of entry for potential entrants that draw the permanent idiosyncratic productivity level Z is:

$$F_Z^e = \int_{\mathcal{S}} \int_{\mathcal{Z}} F(Z,z,0) dG' d\tilde{H}'(z). \tag{14}$$

As in many other studies in the entry literature, we assume a one period to build, so that firms that decide to enter today will, if economically convenient, start producing tomorrow. As a result, the free-entry condition reads as:

$$E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}} \tilde{F}_Z^{e'} \geq c_e, \tag{15}$$

where $\tilde{F}_Z^{e'}$ denotes the average value of entrants calculated on the distribution of the idiosyncratic permanent productivity component Z, as perspective entrants pay the entry cost c^e before drawing Z. Condition (15) holds with equality when the mass of *potential* entrants, N^e, is positive.

The right-hand side of (15) is a time-varying entry cost, characterized by the following functional form:

$$c_e = \psi(N^e)^{\xi}, \tag{16}$$

where ψ is subject to shocks: $\psi = \overline{\psi} + \zeta$, $\zeta' = \rho_A \zeta + \epsilon^A$, with $\epsilon^A \sim N(0, \sigma_A)$. As in Colciago and Silvestrini (2022), and Gutiérrez et al. (2021), c_e is convex ($\xi > 0$) and increasing ($\overline{\psi} > 0$) in the mass of potential entrants N^e. This could be motivated by various factors, among which we can list congestion externalities, as in Jaef and Lopez (2014), or diminishing quality in managerial ability, as in Bergin et al. (2018).

The mass of firms evolves as:

$$N' = \int_{\mathcal{Z}} \int_{\mathcal{S}(Z,z',n)} dG(c^f)\, dH(z'|z)\, d\Lambda_{-1}(Z,n,z) + N^e, \tag{17}$$

where Λ_{-1} is the distribution of operating firms at the beginning of the period (states (Z,n,z)), the inner integral over c^f is taken over the survival region $\mathcal{S}(Z,z',n) = \{c^f : \tilde{F}(Z,z',n) > c^f\}$ (as in (10)), and N^e is the mass of new producers entering operation. Equivalently, the next-period distribution Λ' satisfies, for every measurable set B in the state space,

$$\Lambda'(B) = \int_{\mathcal{Z}} \int_{\mathcal{S}(Z,z',n)} \mathbb{I}\{(Z,n,z') \in B\}\, dG(c^f)\, dH(z'|z)\, d\Lambda_{-1}(Z,n,z) + \Lambda^e(B), \tag{18}$$

where $\mathbb{I}\{\cdot\}$ is an indicator function, and Λ^e is the measure induced by successful entry (so $\Lambda^e(\cdot)$ has total mass N^e). Not all potential entrants or incumbents become producers in the next period; only those that survive their fixed cost draw and receive an idiosyncratic productivity shock will operate.

4.2.2 Retailers and the Final Good Producer

There is a fixed mass of retailers, each of them producing a differentiated variety using the good of a production firm as their only input. Pricing is subject to the Rotemberg (1982) mechanism. A perfectly competitive final good producer takes the differentiated products of retailers and bundles them in a single good using a production function with a constant elasticity of substitution, ϵ . Imposing symmetry across retailers, the retailers and final good producers aggregate into a familiar New Keynesian Phillips Curve:

$$\pi(\pi - 1) = E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\,\pi'(\pi' - 1)\frac{Y'}{Y} + \frac{1 - \epsilon + \epsilon\rho}{\phi}, \tag{19}$$

where π is inflation, Y is output, ϕ is the Rotemberg's (1982) adjustment cost parameter. Note that the price of the production firm's good relative to the price of the final good, i.e., ρ, is the marginal cost of retailers. When aggregate demand for the final good decreases, following for example a contractionary monetary policy shock, retailers would decrease production because of the nominal rigidities. Hence, they decrease demand for the heterogeneous firms' goods, so the relative price, ρ, decreases, generating lower inflation through (19).

4.3 Households

The economy is populated by a representative household consisting of a continuum of individuals who can be either employed or unemployed. The mass of employed workers is denoted with L, while that of the unemployed is U. Members of the household pool their income from work and non-work activities and spend it on consumption goods. The present discounted value of lifetime utility is:

$$\mathcal{U} = u(C) + \beta E\mathcal{U}', \tag{20}$$

where C denotes consumption, and the period utility is concave, so that agents are risk averse. The budget constraint of the household reads:

$$C + B/P = \int_{\Lambda} wn\, d\Lambda + bU + \int_{\Lambda} \Pi\, d\Lambda - c_e N^e + B_{-1}R/P - T, \tag{21}$$

where B is the total investment in a nominal risk-less bond issued by the government and P the aggregate price index, b is unemployment benefit financed through lump sum taxes, T. The measure Λ has total mass $N = \int d\Lambda$ (operating firms). Firm-level dividends, i.e. revenues minus labor, vacancies, and fixed costs, gross of the total entry cost $c_e N^e$ are denoted by Π. The real stochastic discount factor $\tilde{\boldsymbol{\beta}}$ is given by the standard Euler equation (see Appendix D for the derivation of the household's first-order conditions):

$$\frac{E_{\lambda'|\lambda}\pi'}{R} = \beta E_{\lambda'|\lambda}\frac{\zeta'}{\zeta} = \tilde{\boldsymbol{\beta}}.$$

4.4 Wage bargaining

In a frictional labor market, the formation of an employment relationship entails a positive surplus that has to be split between the worker and the firm. Decreasing returns to scale at the firm level imply that the surplus depends on the number of workers employed at a given firm. Bargaining is conducted according to the scheme proposed by Stole and Zwiebel (1996), as in Elsby and Michaels (2013) and Sedláček (2020). This bargaining framework is equivalent to Nash bargaining, but bargaining happens over the marginal surplus.[6] The resulting bargained wage is:

$$w(Z, z, n) = \eta \left[\frac{\alpha \rho Z z n^{\alpha - 1}}{1 - \eta(1 - \alpha)} + \kappa E_{\lambda'|\lambda} \tilde{\boldsymbol{\beta}}\, \theta' \right] + (1 - \eta) b. \tag{22}$$

The wage shares costs and benefits associated with the match according to the extent of the bargaining power, as measured by η. The worker is rewarded for a fraction η of the firm's revenues and savings of hiring costs and compensated for a fraction $1 - \eta$ of the foregone unemployment benefits. Starting with Hall (2005) and Shimer (2005), the literature pointed out that search and matching models account for the cyclical properties of unemployment and vacancies when the real wage does not display sharp swings in response to shocks. This led several authors to augment the search and matching framework with a wage norm that dampens fluctuations in the real wage. We follow Sedláček (2020) and model real wage rigidity in the form of a social norm where wages in individual firms are a weighted average of the Stole-Zwiebel wage in equation (22) and its steady-state counterpart. The weight λ_w given to the steady state wage determines the degree of wage rigidity. Notice that $\lambda_w = 0$ corresponds to the case of flexible wages.

4.5 Monetary Policy

The monetary authority sets the nominal risk-free interest rate, R according to a standard Taylor rule:

$$R = \tilde{R}_{-1}^{\phi} \pi^{\phi(1-\phi^R)} Y^{\phi(1-\phi^R)} e^{\epsilon^R}, \tag{23}$$

where ϵ_t^R is a normally distributed, zero mean, monetary policy shock.

[6] See Appendix E for the derivation of the wage equation (22).

4.6 Aggregation, market clearing, and equilibrium

The resource constraint of the economy can be written as:

$$C = Y - \kappa V - N\overline{c_f} - c_e N^e, \tag{24}$$

where $Y = \int_\Lambda y \, d\Lambda$ is aggregate output, $\overline{c_f}$ is an average of the fixed costs drawn by continuing businesses based on the firm size distribution, and $V = \int_\Lambda v \, d\Lambda$ are total vacancies. Equation (24) states that aggregate output is spent on consumption, the cost of creating vacancies, fixed costs of production, and the cost of creating new firms. Notice that GDP equals aggregate output net of fixed and vacancy creation costs. The following conditions define the model's equilibrium. The mass of employed members of the representative household equals the mass of workers employed at operating firms; the employment policy function of individual firms is determined according to (11); wages are determined according to the weighted average between the wage defined in equation (22) and its steady state counterpart. Finally, the free-entry condition (15) holds with equality and the total mass of operative firms evolves according to (17).

5 Parameterization, Estimation, and Solution Method

Since the model does not admit a closed-form solution, we solve it numerically to compute both the stationary equilibrium and the dynamic response to a monetary policy shock. A subset of parameters is assigned values based on external empirical evidence, another subset is calibrated to match key steady state moments, and the remaining parameters are estimated by matching model-implied impulse responses to their empirical counterparts.

We compute the stationary equilibrium by applying a local approximation of the value function. The full solution algorithm is presented in Appendix F. To obtain the dynamic stochastic equilibrium, we use the method developed by Reiter (2009).[7]

Calibrated parameters. The time period is a quarter. Preference and technology parameters follow standard values in the literature. The discount factor is set to $\beta = 0.99$, and the parameter measuring the degree of returns to scale is set to $\alpha = 0.65$. Turning

[7]We use the MATLAB code provided by Costain and Nakov (2011), available on their web pages.

to the parameters belonging to the search and matching framework, we choose the mass of workers, employed and unemployed, such that the steady state labor market tightness θ is normalized to one. The matching elasticity is set to $\gamma = 0.72$, following Shimer (2005). Workers' bargaining power η is equal to 0.5, which maintains comparability with much of the search and matching literature.[8] The matching efficiency parameter μ is chosen to match a job finding probability equal to 0.7, in line with Blanchard and Galí (2010a). Turning to the parameters that characterize the entry cost process, the elasticity of entry costs with respect to the mass of potential entrants ξ is set to 1.5, as estimated by Gutiérrez et al. (2021). The parameter ψ determines the mass of potential entrants N^e through the free entry condition and therefore affects the overall scale of the economy, but not the moments that we target below. We therefore normalize its value to one. We set the degree of wage stickiness λ_w equal to 0.7 as a starting point for the estimation stage.

Parameters and moments of the stationary distribution. We jointly calibrate some parameters to minimize the Euclidean distance between model-implied moments in the stationary equilibrium and their empirical counterparts. The target moments are firm shares, employment shares, and exit rates across three firm size groups, namely new entrants, small firms, and large firms, constructed from the reorganized Business Dynamics Statistics data. In addition, we target an aggregate separation rate equal to 9 percent. This value lies in the range reported by Hall (2005), who document separation rates between 8 and 10 percent for the U.S. economy. Finally, we target a ratio between total vacancy posting costs and GDP, $\frac{\kappa V}{Y} = 0.01$, consistent with standard calibrations of hiring and recruiting costs in the U.S. labor market, as in Blanchard and Galí (2010b).

The calibrated parameters include the mean and dispersion of fixed operating costs, μ_o and σ_o, the standard deviation σ_z and persistence ρ_z of idiosyncratic productivity shocks, the distribution G of permanent idiosyncratic productivity, the vacancy posting cost κ, and the unemployment benefit b. We adopt a flexible specification for the distribution G of permanent idiosyncratic productivity. Following Sedláček (2020), we assume two productivity types, ordered by efficiency, each drawn with an endogenous probability. Together with the calibration of ρ_z, this structure allows the model to closely match the joint distribution

[8] As noted by Gertler and Trigari (2009), this value also lies within the range estimated by Flinn (2006).

Parameter or SS value	Definition	Value	Source or Target
β	Discount factor	0.99	Interest rate equal to 4 percent
α	Returns to scale parameter	0.65	Labor share
ϵ	Elasticity of substitution	3.8	Ghironi and Melitz (2005)
θ	Labor market tightness	1	Shimer (2005)
γ	Matching elasticity	0.72	Shimer (2005)
η	Workers' bargaining power	0.5	Flinn (2006)
ξ	Elasticity of entrants	1.5	Gutiérrez et al. (2019)
ψ	Entry cost parameter	1	Normalization
μ_o	Log normal parameter	-7.59	BDS distributions
σ_o	Log normal parameter	4.54	BDS distributions
Z_j	Permanent productivity types	[1.94, 11.001]	BDS distributions
$F(Z_j)$	Probability of each type	$[0.9999, 10^{-4}]$	BDS distributions
ρ_z	Persistence of idiosyncratic shocks	0.87	BDS distributions
σ_z	Standard deviation of idiosyncratic shocks	0.11	Separation rate equal to 9 percent

Table 2: Calibrated parameters

of firm size and exit rates observed in the data. As shown in the counterfactual analysis, matching exit rates across firm sizes is crucial for explaining the response of employment concentration to a monetary policy shock. Under this calibration, the model implied aggregate quarterly exit rate in the stationary equilibrium is close to 2.5 percent. This value is in line with the exit rates commonly assumed in firm dynamics models with exogenous exit, such as Bilbiie et al. (2012).

Parameters estimated by impulse response function matching. We next estimate the structural parameters that do not affect the steady state, together with two key parameters that do influence the stationary distribution of the model: the vacancy posting cost parameter, κ, and the unemployment benefit, b. Unlike most of the remaining parameters, which primarily govern the model's dynamic responses, κ and b play a first-order role in shaping the stationary equilibrium. In particular, they jointly influence job creation incentives, equilibrium unemployment, separation rates, and the stationary distribution of firms and employment across size classes, through entry and exit. Moreover, direct empirical evidence on their values is limited.[9]

The estimation is conducted using a limited information minimum distance estimator,

[9]The values of κ and b calibrated in the previous paragraph are taken as initial values in the estimation routine.

following Trigari (2009). This approach builds on the impulse response matching strategies developed by Rotemberg (1997), Christiano et al. (2005), and Boivin and Giannoni (2006). The estimator selects parameter values that minimize the distance between a set of model-implied impulse responses and their empirical counterparts obtained from our Bayesian VAR after a monetary policy shock.[10]

While this approach is well suited to disciplining the model's dynamic responses, parameter values that improve the fit of the model's responses to a monetary policy shock may simultaneously generate counterfactual stationary distributions. Specifically, matching impulse responses do not discipline κ and b with respect to their implications for firm shares, employment shares, exit rates, and the aggregate separation rate.[11]

To complement impulse-response matching with information from the stationary equilibrium, we augment the limited-information minimum distance objective with a penalty function, $F^P(\cdot)$, which penalizes deviations of selected stationary moments—namely firm shares, employment shares, exit rate by firm size, and aggregate separation rate—from their empirical counterparts.

The resulting augmented minimum distance criterion is defined as follows. Given the penalty function $F^P()$, let Ω^{DSGE} denote the vector of estimated structural parameters, and let $\Gamma(\Omega^{DSGE})$ collect the model-implied impulse responses to a monetary policy shock. Let Γ_{BVAR} denote the vector of median BVAR impulse responses to the same shock.

The minimum distance estimator $\tilde{\Omega}^{DSGE}$ solves

$$\begin{aligned}\min_{\Omega^{DSGE}} L(\Omega^{DSGE}) = &\left[\Gamma(\Omega^{DSGE}) - \Gamma_{BVAR}\right]' \mathbf{W} \left[\Gamma(\Omega^{DSGE}) - \Gamma_{BVAR}\right] \\ &+ F^P\Big(|\boldsymbol{\Delta M}|\Big),\end{aligned} \tag{25}$$

where $\mathbf{W}$ is a diagonal weighting matrix with diagonal elements equal to the inverse of the posterior variance of each BVAR impulse response. The penalty function $F^P(\cdot)$ enforces consistency between selected steady-state moments of the model and their empirical counterparts. Its argument $|\boldsymbol{\Delta M}|$ is the vector of absolute deviations between model-implied

[10] We focus on impulse responses for variables that are central to our analysis of the labor market and to the transmission of monetary policy, namely employment concentration, unemployment, inflation, GDP, and the nominal interest rate.

[11] Elsby and Michaels (2013) show that the unemployment benefit interacts with the dispersion of idiosyncratic shocks to affect equilibrium separation rates.

and data moments, including employment shares, firm shares, exit rates by firm size, and the aggregate separation rate.[12]

Parameter	Description	Value (Std. Error)
κ	Cost of posting a vacancy	0.408 (0.00002)
b	Unemployment benefit	1.303 (0.00005)
λ^w	Degree of wage stickiness	0.9753 (0.00083)
σ^R	Standard deviation of monetary policy shocks	0.122 (0.0543)
b^C	Strength of consumption habits	0.935 (0.00008)
ϕ^R	Interest rate smoothing parameter	0.731 (0.2727)
ϕ_π	Policy rule parameter on inflation	1.513 (4.03)
ϕ_y	Policy rule parameter on the output gap	0.0131 (0.57)
ρ^R	Persistence of monetary policy shocks	0.854 (0.125)
ψ^R	Rotemberg price adjustment parameter	101.986 (13.2)

Table 3: Estimation results

During the estimation, the algorithm updates the coefficient governing the Rotemberg (1982) price adjustment costs, denoted by ψ^R, to target a slope of the Phillips curve equal to 0.1, following Ottonello and Winberry (2020). The estimated vacancy posting cost implies that total vacancy posting costs amount to about 0.5 percent of GDP in the stationary equilibrium, somewhat below the 1 percent target used in the calibration. To assess the plausibility of this estimate, we examine the implied ratio between the vacancy posting cost and the average wage. In the stationary distribution, the estimated value of κ corresponds to a vacancy posting cost equal to roughly one quarter of the average wage, a magnitude well within the range commonly used in search and matching models of the U.S. labor market (see, among others, Hagedorn and Manovskii 2008).

The estimated unemployment benefit implies a replacement rate of approximately 70 percent of the average wage. While this value is on the high side relative to some calibrations, it allows the model to generate a stationary separation rate of 9.7 percent, given the calibrated value of the dispersion of idiosyncratic productivity shocks.[13] This value is consistent with empirical estimates reported by Hall (2005) and close to the 10 percent separation rate commonly used in the literature, including Shimer (2005).

[12] A detailed description of the penalty function is reported in Appendix G.

[13] Elsby and Michaels (2013) show that b and the dispersion of idiosyncratic productivity shocks, σ_z, jointly determine the separation rate.

	Targeted Moments			Untargeted JCR and JDR moments		
Firm Category	**Firm Share**	**Emp. Share**	**Exit Rates**	**JCR**	**JDR**	**JDR^{exit}**
Model						
New (Entrant)	0.0252	0.0096	0	2	0	0
Small (1-499)	0.9688	0.4969	0.0261	0.0528	0.0765	0.0254
Large (500 or more)	0.0060	0.4936	4.16e-04	0.0869	0.0991	4.11e-04
BDS data (quarterly)						
New (Entrant)	0.0263	0.0067	0	2	0	0
Small (1-499)	0.9694	0.4990	0.0232	0.0344	0.0408	0.0115
Large (500 or more)	0.0042	0.4943	4.5634e-04	0.0318	0.0301	5e-04

Table 4: Long-run distributions: model vs data.

6 Stationary distributions

In Table 4 we report the model-implied firm shares, employment shares, and exit rates across new entrants, small firms, and large firms, which were the targets of our calibration strategy, and compare them to those extracted from BDS. The model matches the targeted moments well. A key feature of the data is that large firms exit with a much lower frequency from the market than small firms. In the remainder, we demonstrate that matching this aspect of the data not only enhances the empirical appeal of the model, but is also crucial for explaining the response of employment concentration to monetary policy shocks.

The table also reports the model-implied JCR and JDR for entering firms and for continuing ones according to size, together with JDR due to the exit of firms from the market.

When computing job flow rates by firm size, we follow Davis et al. (1996) and use, in the denominator, the average employment across periods. Consider job creation in a specific category s, where $s \in \{\text{new, small, large}\}$.[14] Let $E_{i,t}$ denote the employment of firm i at time t, and define the firm-level average employment as:

$$\bar{E}_{i,t} = \frac{E_{i,t-1} + E_{i,t}}{2}.$$

[14] Averaging employment over two adjacent periods mitigates regression-to-the-mean bias, which arises when firms are classified by size at a single point in time and may misrepresent job flows due to transitory shocks. See Davis et al. (1996) for details.

The average employment for the category is then obtained by summing over all firms $i \in s$. The job creation rate for category s is defined as:

$$JCR^s = \frac{\sum_{i \in s} \max(E_{i,t} - E_{i,t-1}, 0)}{\sum_{i \in s} \bar{E}_{i,t}},$$

where the numerator sums only positive employment changes across all firms in the category. By symmetry, the job destruction rate is:

$$JDR^s = \frac{\sum_{i \in s} \max(E_{i,t-1} - E_{i,t}, 0)}{\sum_{i \in s} \bar{E}_{i,t}}.$$

This approach ensures that both rates are measured relative to the same average employment base, making them comparable across size categories and periods. The job destruction rate due to firm exit in category s can be defined by restricting the numerator to the employment of firms that exit between $t-1$ and t:

$$JD^{s,\text{exit}} = \frac{\sum_{i \in s,\text{exit}} E_{i,t-1}}{\sum_{i \in s} \bar{E}_{i,t}}.$$

This decomposition allows us to separate total job destruction into contributions from exiting firms and from continuing firms that reduce employment.

Although these distributions are not specifically targeted in our calibration strategy, the model matches them quite well. Specifically, the model matches the fact that the JDR due to exit by large firms is close to zero. This will play a pivotal role at improving the matching of the empirical trade-off between inflation and unemployment in response to policy shocks.

7 Dynamics: Monetary Policy Shock

In this section, we study the transmission of monetary policy shocks in our model.

Panels (a) to (e) of Figure 4 report the impulse response functions targeted in the estimation: employment concentration, unemployment, GDP, inflation, and the nominal interest rate. The black solid lines show the median responses from the BVAR, while the blue dashed lines report the model-implied responses. Panel (f) displays the corresponding Phillips multiplier. Shaded areas denote 68 percent credible intervals.

The model replicates the key empirical patterns. Following a contractionary monetary

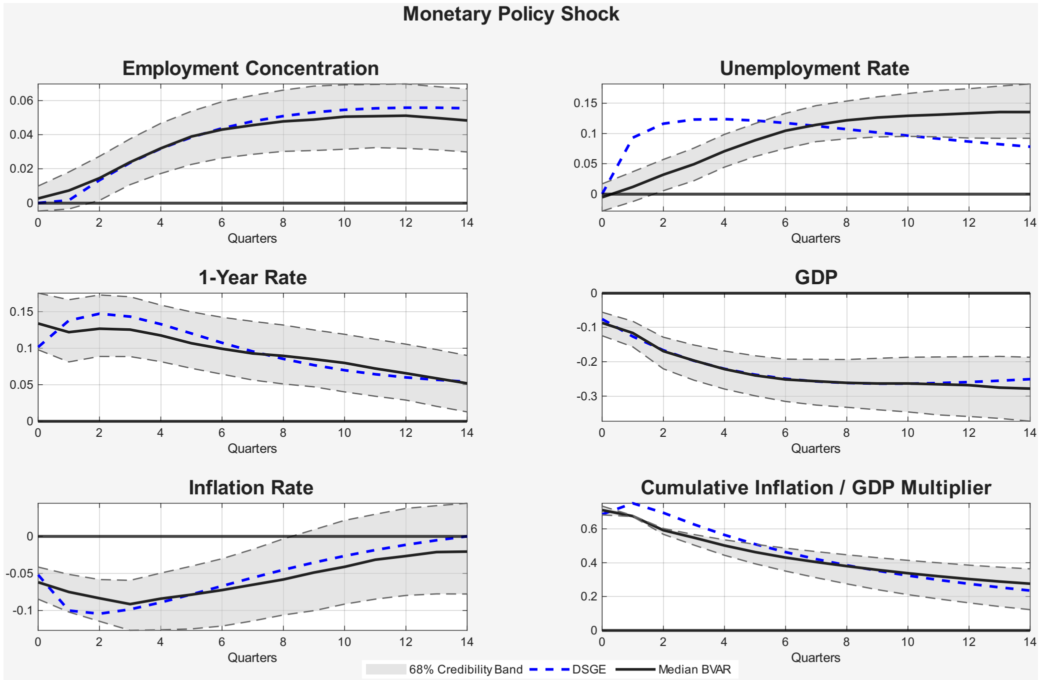

Figure 4: Impulse response functions matching.

policy shock, employment concentration increases gradually and persistently, unemployment rises, and both output and inflation decline. The output response is highly persistent, while inflation reaches its trough after a few quarters and then gradually recovers. The model also matches closely the empirical Phillips multiplier.

The increase in employment concentration is driven by firm-level heterogeneity in entry, exit, and job creation. Small firms exhibit substantially higher exit rates than large incumbents, while new entrants account for a disproportionate fraction of job creation compared to their employment share. A contractionary monetary policy shock reduces households' investment in new firms, leading to a sharp decline in entry and in job creation by entrants, which are predominantly small. At the same time, the higher interest rate lowers demand and increases firm exit, disproportionately affecting small firms. As a result, the contraction disproportionately affects small and young firms, while large incumbents account for a larger share of the remaining activity. This leads to a persistent shift in the distribution of employment toward large firms and, therefore, an increase in employment concentration. The close match between the model-based and empirical impulse responses indicates that this employment concentration channel is a key mechanism through which monetary policy

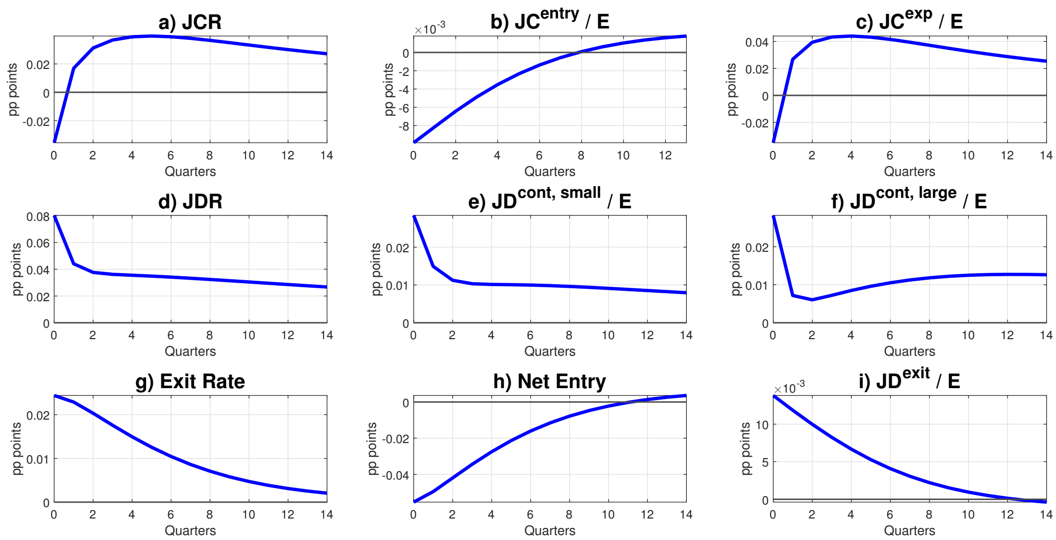

Figure 5: IRFs of Job flows and other key variables to a MP shock.

affects employment concentration in the data.

Figure 5 shows the model counterpart of Figure 2 in the empirical analysis. It further confirms the empirical consistency of the model's transmission mechanism. In response to the policy shock, job creation by new entrants declines, while job creation by incumbent firms initially falls but subsequently rises. This rebound in incumbents drives the overall dynamics of total job creation. Total job destruction increases following the shock (panel d). Consistently with the data, the model also features an increase in the exit rate and in job destruction due to exit. Since large firms essentially never exit, this margin is entirely driven by small firms. The joint decline in entry and rise in exit among small firms therefore affects both the amount of productive capacity lost and the composition of firms accounting for the remaining production. This distinction is central for aggregate dynamics. On the one hand, higher exit reduces the mass of active firms, lowering effective productive capacity. On the other hand, the fact that exit is concentrated among small firms changes the composition of production toward large incumbents. These two forces jointly determine how the decline in demand affects aggregate dynamics, and therefore shape the model-implied output–inflation trade-off. In the next section, we isolate these mechanisms through counterfactual exercises to highlight which features are essential for generating the observed responses.

8 Counterfactual Analysis

The central premise of our analysis is an empirically grounded distinction between large and small firms, capturing both their differential exposure to entry and exit dynamics and their contribution to aggregate employment. In this section, we clarify why these features are crucial for shaping the aggregate response to a monetary policy shock.

The key point is that aggregate outcomes depend on how monetary policy affects productive capacity through its impact on firms' entry and exit decisions. Following a contractionary monetary policy shock, demand declines and some firms exit, reducing the mass of active producers. The aggregate implications of this adjustment depend on two distinct elements: the strength of the extensive margin, i.e., how many firms enter and exit, and its composition, i.e., which firms enter and exit. Crucially, what matters for equilibrium prices is how these margins reshape the contribution of active firms to aggregate supply.

A useful way to organize the mechanism is thus around two distinct effects. The first is a *shift effect*: following a contractionary monetary policy shock, the rise in exit reduces the mass of active firms and thereby shifts aggregate supply inward. The second is a *slope effect*: the response of prices and quantities depends not only on how many firms exit, but also on which ones, that is, on how the shock changes the composition of active firms and hence the slope of the aggregate supply curve. This slope reflects how production and employment are distributed across firms, while its sensitivity to the shock is governed by selection and cleansing along the entry and exit margins.

These two forces are conceptually distinct. The shift effect captures the direct contraction in supply resulting from a decline in the mass of active producers. The slope effect, instead, captures how a given change in the mass of firms maps into aggregate outcomes depending on which firms exit and how production is distributed across continuing firms.

To isolate these mechanisms, we consider two counterfactual exercises. In the first, we shut down endogenous exit by imposing a constant, size-independent exit probability, so that exit does not respond to the monetary policy shock. This counterfactual mainly focuses on the shift effect, because it suppresses the additional contraction in supply generated by the endogenous rise in exit characterizing the baseline framework. In the second, we vary

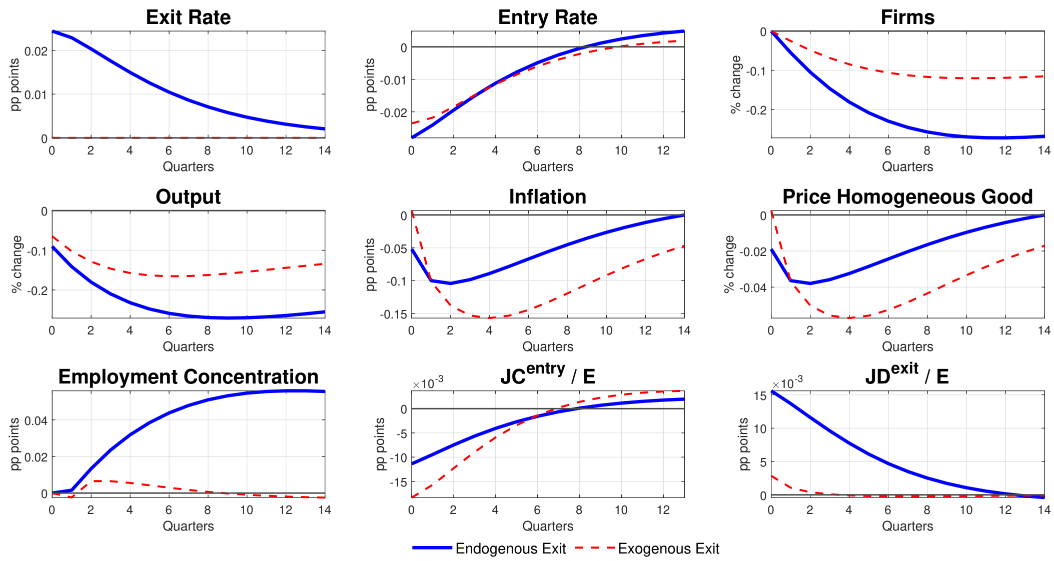

Figure 6: IRFs under exogenous and endogenous exit.

the degree of employment concentration in the stationary distribution. This counterfactual mainly speaks to the slope effect, because it changes how a given pattern of exit and entry maps into the composition of production and thus into the slope of aggregate supply. When production is concentrated in a small number of highly productive firms, the exit of smaller firms has limited effects on aggregate capacity. When, instead, production is more evenly distributed, exiting firms account for a larger share of output, so changes in the composition of firms translate into a more muted response of prices and a sharper contraction in supply.

Comparing impulse responses across these economies allows us to separately assess how monetary policy transmission depends, first, on the extent to which the shock changes the mass of active firms and, second, on how strongly those changes reshape the distribution of production across them. Through this lens, heterogeneity matters not per se, but because it governs both the size of the endogenous supply contraction and the sensitivity of the aggregate supply schedule to demand-driven selection and cleansing.

8.1 Exogenous Exit

We first shut down endogenous exit by imposing a constant, size-independent exit probability equal to the average exit rate in the baseline. Figure 6 compares the impulse responses under this specification (red-dashed lines) with those in the baseline economy (blue-solid lines).

By construction, the exit rate does not respond to the shock in this counterfactual (panel a), whereas it increases in the baseline. This translates into a muted response of job destruction though exit (panel i). Since entry responds similarly across specifications (panel b), the mass of firms declines by less (panel c), resulting in a smaller contraction in output (panel d).

This difference isolates the *shift effect*: when exit does not rise endogenously after the monetary contraction, the reduction in the number of active producers is smaller, and so is the inward shift in aggregate supply.

The price response moves in the opposite direction. In the baseline, the larger increase in exit shifts aggregate supply inward and thereby attenuates the decline in prices. Under exogenous exit, this supply contraction is muted. As a result, the decline in demand is absorbed to a greater extent through prices: inflation falls more and remains below baseline for longer (panel e), and the price of the homogeneous good also declines more strongly (panel f).

Imposing size-independent exit also weakens the concentration channel, which governs the slope effect. Because exit no longer responds differentially across firm sizes, employment concentration reacts only modestly (panel g), in contrast with the baseline where the contraction of smaller firms induces more concentration toward larger incumbents. This composition channel reinforces the difference in price adjustment.

A useful way to summarize the mechanism is through the slope of the aggregate supply curve and its sensitivity to the shock. The overall decline in the mass of firms shifts aggregate supply inward, i.e., the shift effect, while changes in the distribution of production across active firms, i.e., the concentration channel, alter its slope, i.e., the slope effect. While this counterfactual focuses mainly on the former effect, the next one leaves the former largely unchanged, but changes the latter substantially.[15]

[15] Appendix H presents a simple model of a competitive market with heterogeneous firms, quadratic cost functions, and linear demand. In that setting, aggregate supply can be written as $Q = S(P) = PH(P)$. The object $H(P)$ summarizes how the composition of active firms shapes the aggregate supply schedule. It depends both on the mass of active firms and on the distribution of production shares across them. A demand contraction affects equilibrium not only directly, but also indirectly through its effect on $H(P)$: endogenous entry and exit change both the number of active firms and the distribution of production across them, thereby shifting aggregate supply and altering its slope. The extent to which the shock is absorbed by prices rather than quantities depends on how strongly entry and exit affect these two margins.

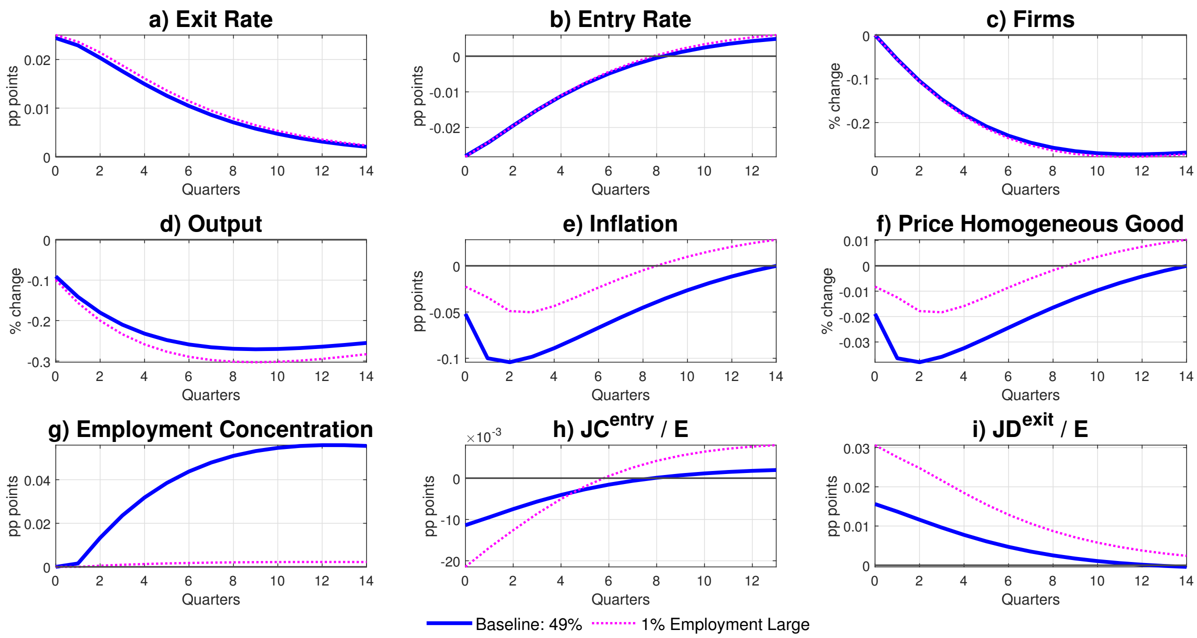

Figure 7: IRFs under alternative degrees of employment concentration.

8.2 Employment Concentration

We next vary the degree of employment concentration by considering an alternative stationary distribution in which large firms account for only 1 percent of total employment. This configuration approximates an economy in which production is more evenly distributed across firms.

Figure 7 reports the impulse responses under this low-concentration economy (red-dotted lines) and in the baseline (blue-solid lines). A first striking feature is that entry and exit behave very similarly across the two economies (panels a–b), implying only modest differences in the change in the mass of of active firms (panel c). Relative to the previous counterfactual, therefore, this exercise does not primarily operate through a different response of the extensive margin. The key difference is instead compositional: given broadly similar firm dynamics, the same amount of exit has different aggregate consequences across specifications because it removes firms that account for different shares of production and employment.

Two related mechanisms are at work, related to the extensive and intensive margin, respectively. First, in the low-concentration economy, production and employment are distributed more evenly across firms. As a result, the firms that exit after the shock account for a larger share of aggregate activity than in the baseline. Thus, even if the increase in exit is similar across the two economies, a given amount of turnover induces a larger inward shift in aggregate supply in the low-concentration case. In the baseline economy, by contrast, production is concentrated in a small number of large and highly productive firms. Exit is again concentrated among smaller firms, but these firms account for only a limited share of aggregate production and employment. For a given increase in exit, aggregate supply therefore contracts by less.

Second, regarding the intensive margin, the effects of the shock are distributed differently across surviving firms. This is where employment concentration becomes central: its response (panel g) is the clearest manifestation of the mechanism in the model. In the low-concentration economy, the decline in demand is absorbed more evenly across incumbents, so employment concentration changes only modestly after the shock. In the baseline, by contrast, the contraction is borne disproportionately by a narrower set of extant large firms, and employment concentration rises more strongly. The response of average firm size is consistent with this interpretation: in the low-concentration economy, the average size of small firms declines by less than in the baseline, indicating that the contraction is spread more broadly across surviving firms.

With increasing marginal costs, this difference in the intensive-margin adjustment has direct implications for prices. When the contraction is distributed more evenly across many relatively similar firms, each surviving firm reduces production by less, so firm-level marginal costs decline by less and the price response is more muted. When instead the contraction is concentrated on a smaller set of large firms, their production adjusts more sharply, their marginal costs fall by more, and prices respond more strongly. In this sense, the slope effect in this counterfactual operates through the response of employment concentration: it captures how a given decline in demand is allocated across surviving firms, and therefore how strongly marginal costs and prices adjust.

This second counterfactual therefore differs from the exogenous-exit exercise in an important respect. In the previous counterfactual, the main mechanism operated through the extensive margin: shutting down the endogenous response of exit weakened the contraction in the mass of active firms and thereby altered the aggregate response of prices and quantities. Here, instead, entry and exit rates by firm type remain very close to the baseline. The main difference therefore does not arise from a materially different turnover response, but from differences in the aggregate importance of the firms that exit and in how the contraction is reallocated across surviving incumbents. While the first counterfactual mainly isolates a shift effect operating through endogenous exit, this second counterfactual primarily isolates a slope effect operating through employment concentration and the associated response of marginal costs.

9 Concentration, Exit, and the Trade-Off Between Output and Inflation

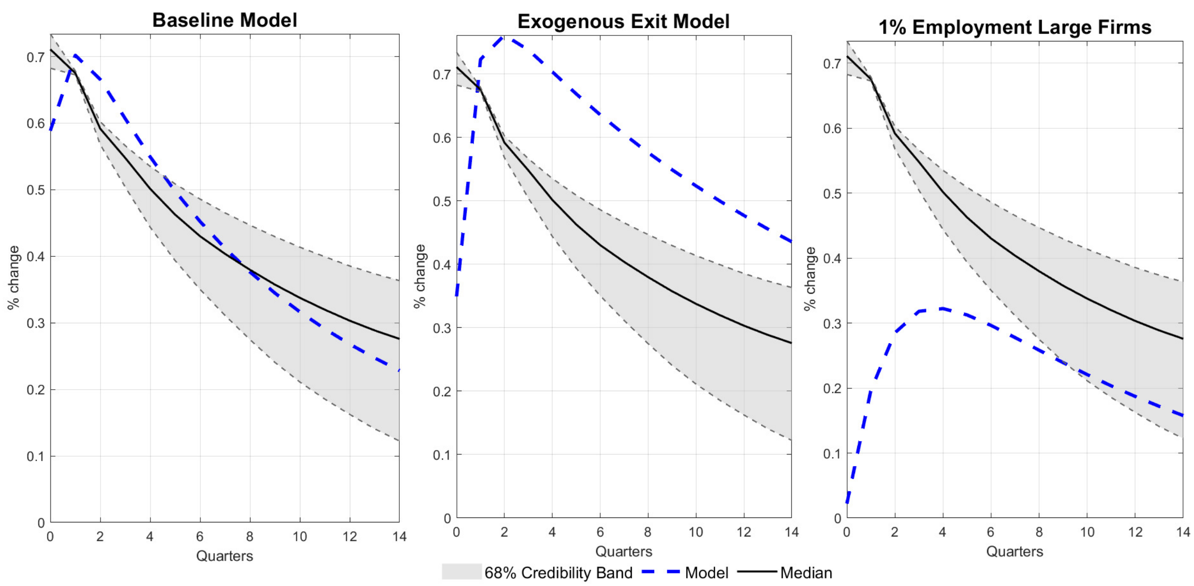

Figure 8: Phillips multipliers across estimated model specifications.

The counterfactual analysis in the previous section shows that both the degree of employment concentration and the endogeneity of the exit margin play a central role in the transmission of monetary policy shocks. The relative responses of output and inflation, i.e., the output–inflation trade-off monetary policy is facing, thus depend critically on firm composition.

We now evaluate how these two dimensions jointly shape the trade-off between output and inflation, as summarized by the Phillips multiplier. To ensure a consistent comparison across model specifications, for each restricted model, we re-estimate the relevant set of structural parameters using the same limited-information minimum distance procedure described in Section 5. Table 5 in Appendix I reports the estimated parameters across all three specifications. The estimates are stable: all parameters remain virtually unchanged across models, with the most notable variation appearing in the Rotemberg price adjustment parameter ψ_R, which rises from 102.0 in the baseline to 104.4 in the low concentration economy, and to 115.6 in the economy with exogenous exit. Thus, the model requires a higher degree of price rigidities to match the empirical IRFs, when there are few small fragile firms, and particularly so when exit rates are identical across firms.[16]

Figure 8 reports cumulative Phillips multipliers across the specifications we considered. The left panel shows the baseline, while the middle and right panels refer to the two restricted specifications. In each panel, the model-implied multiplier (blue-dashed line) is compared to its empirical counterpart, depicted with a black-solid line, together with the associated 68 percent credible interval.

The baseline model, with empirically realistic employment concentration and endogenous, size-dependent exit, closely matches the empirical Phillips multiplier. Both the magnitude and the dynamic profile of the output–inflation trade-off implied by the model align well with the data, indicating that the baseline model is able to capture the key mechanisms governing the aggregate response to monetary policy tightening.

In contrast, deviations from this configuration generate output–inflation trade-offs that are at odds with the empirical evidence.

When suppressing endogenous, size-dependent exit, the multiplier is overstated (except in the initial 2 periods), due to combination of limited output contraction and the more persistent decline in inflation discussed in the previous section. Conversely, when employment concentration is very low, the economy behaves similarly to a homogeneous-firm environment, and the model substantially understates the Phillips multiplier, reflecting excessive output losses relative to changes in inflation. Both restricted models, therefore, cannot

[16] Note however that according to the baseline estimate, these differences are not statistically significant. Moreover, Appendix I reports the counterparts of the impulse responses in Figures 6 and 7 using parameters estimated for the restricted models (see Figures A3 and A4). Since the responses are essentially unchanged, the differences reported in Sections 8.1 and 8.2 reflect the imposed structural restrictions rather than parameter re-estimation.

replicate the empirical Phillips multiplier.

The counterfactual analysis highlights that the output–inflation trade-off depends critically on firm composition. When exit rates are constant and identical across firms, the absence of destruction of production capacity due to endogenous exit leads to a strong reaction of inflation and an overstatement of the multiplier. In a low concentration economy dominated by small fragile firms, job destruction is widespread and the central bank faces a lower trade-off between output and inflation. By contrast, when a realistic share of employment is concentrated in large incumbents, which are insulated from entry and exit margins, the concentration channel amplifies the effect on inflation, leading to greater disinflation for a given output change and, therefore, a steeper trade-off.

These results indicate that reconciling model-implied and empirical Phillips multipliers requires a combination of realistic employment concentration and endogenous, size-dependent exit. Since estimated parameters remain nearly identical across specifications, this conclusion follows from the structural mechanisms of the model, that govern how monetary policy affects firm dynamics, job destruction, and labor concentration, and are essential to reproduce the output–inflation trade-off observed in the data.

10 Conclusions

In this paper, we show that contractionary monetary policy reshapes the economy through firm composition, that is, through the distribution of employment between fragile, high-churn small firms and large, established incumbents. Monetary tightening affects these groups asymmetrically. It suppresses firm entry and increases the exit of small firms, disrupting the flow of job creation that sustains the small-firm segment, while large incumbents remain largely insulated from both margins. As a result, the relative employment share of large firms rises and employment concentration increases persistently, well beyond the initial shock.

To rationalize these findings, we develop and estimate a structural model with heterogeneous firms in a search-and-matching environment, in which firm size is endogenously bounded by diminishing marginal returns to labor. The model features two groups of firms: fragile small firms, which draw lower permanent productivity and operate close to their exit threshold, and large incumbents, which are far from that threshold and rarely exit. The model is disciplined using moments from the Business Dynamics Statistics on firm shares, employment shares, and size-dependent exit rates, and by matching its dynamic responses to monetary policy shocks to their empirical counterparts from the BVAR. This joint discipline ensures that both the stationary distribution of firms and the dynamic transmission of monetary policy are tightly anchored to the data.

Consistent with the evidence, the estimated model replicates the persistent rise in employment concentration following a monetary tightening, the asymmetric dynamics of job creation and destruction across firm sizes, and the empirical output–inflation trade-off. In particular, disinflation is achieved only at the cost of a highly persistent output contraction, and this trade-off deteriorates over the horizon.

Counterfactual exercises isolate the two distinct channels through which firm dynamics shape the aggregate response: a size effect, operating through changes in the mass of active firms, and a slope effect, operating through the composition of production. In the first counterfactual, we shut down endogenous exit. As a result, the contractionary shock destroys less productive capacity, the decline in the mass of firms is muted, and a larger share of the adjustment occurs through prices rather than quantities. In the second counterfactual, we instead vary employment concentration while leaving entry and exit dynamics broadly unchanged. In this case, the key difference is compositional: when production is more evenly distributed across firms, the same amount of exit removes a larger share of effective productive capacity and the contraction is absorbed more through quantities; when production is concentrated in a few large firms, exit affects mostly marginal producers, so capacity changes little and prices adjust more.

Taken together, our findings imply that the output–inflation trade-off is not a primitive of the economy, and it depends jointly on how much productive capacity is destroyed after the shock and on which firms account for that capacity. It is, therefore, endogenous to firm composition and to the asymmetry in how different firms respond to monetary tightening. In economies where employment is concentrated in large incumbents that are insulated from entry and exit margins, a given contraction in demand translates more into price adjustment and less into changes in effective productive capacity. Conversely, when employment is more evenly distributed across firms, the contraction induces a stronger contraction in capacity, and a greater share of the adjustment occurs through quantities rather than prices.

This conclusion has direct and underappreciated policy relevance. In the United States, employment concentration has increased markedly since the 1980s, alongside a secular decline in business dynamism and firm entry. Our framework implies that this structural transformation has not been neutral for monetary transmission. As the economy has shifted toward large, established firms, the mechanism through which monetary policy affects effective productive capacity has changed, and with it the output–inflation trade-off faced by the Federal Reserve. Understanding how changes in market structure interact with monetary transmission is therefore central to assessing whether monetary policy frameworks developed in an era of greater business dynamism remain appropriate for today's more concentrated economy. We leave a systematic quantitative exploration of this question for future research.

Appendix

A Data treatment

Variables from BDS are built as follows:

- We assign firms to three categories: {new, small, large}. New entrants belong to **new** and are classified as firms with size zero. Small firms, with fewer than 500 employees, belong to **small**, while firms with 500 or more employees belong to **large**.[17]

- Small and large incumbent firms are classified using their average size over two adjacent periods to mitigate regression-to-the-mean bias, as discussed in Davis and Haltiwanger (1992).

- Gross job flows (creation of new jobs by incumbents and entrants, and destruction of existing jobs due to either contracting firms or exit) are classified according to the three categories: entrants, small, large.

[17] Large firms are defined as those with 500 or more employees, following Moscarini and Postel-Vinay (2012) and Fort et al. (2013). These references use establishment-level BDS data and adopt a 250-employee cutoff. At the firm level, BDS reports only broader size classes (100–499 and 500 or more), so finer cutoffs cannot be applied.

- Entry and exit rates are defined as the number of firm births and deaths, respectively, relative to the total number of firms operating in the current year.

- Employment concentration measures the distribution of employment between small and large firms. It is defined as:

$$\frac{\text{employment at large firms}}{\text{employment at all other firms}}.$$

B Proxy Structural VAR

This section presents an alternative identification strategy of monetary policy shocks, based on the Proxy Structural VAR approach, following Gertler and Karadi (2015) and subsequent contributions by Miranda-Agrippino (2016), Jarociński and Karadi (2020), and Miranda-Agrippino and Ricco (2021).

We identify monetary policy shocks using the high-frequency external instrument proposed by Miranda-Agrippino (2016), constructed from orthogonalized surprises in federal funds futures.[18] The instrument is plausibly exogenous, independent of central bank forecasts, and unpredictable given past information. Identification is implemented in a monthly VAR that includes the 1-year U.S. Treasury yield (GS1), industrial production, the consumer price index (CPI), the excess bond premium, and the unemployment rate, all sourced from FRED. Using a monthly specification allows us to exploit the external instrument at its original frequency, thereby preserving its informational content and avoiding the loss of identification power that may arise from temporal aggregation.

The resulting impulse responses are reported in Figure A1.

To assess the robustness of our main findings on firm dynamics—specifically entry, employment concentration, exit of incumbents, and job destruction due to exit—we adopt a two-step procedure. First, we recover quarterly series for the BDS variables using Gibbs sampling. This step allows us to reconcile the annual frequency of the BDS data with the quarterly frequency required for the analysis of firm-level outcomes.

Second, we aggregate the identified monthly monetary policy shocks to quarterly fre-

[18]In Miranda-Agrippino (2016), this instrument is denoted by the orthogonal monthly surprise $FF4^*$ derived from the fourth federal funds futures contract.

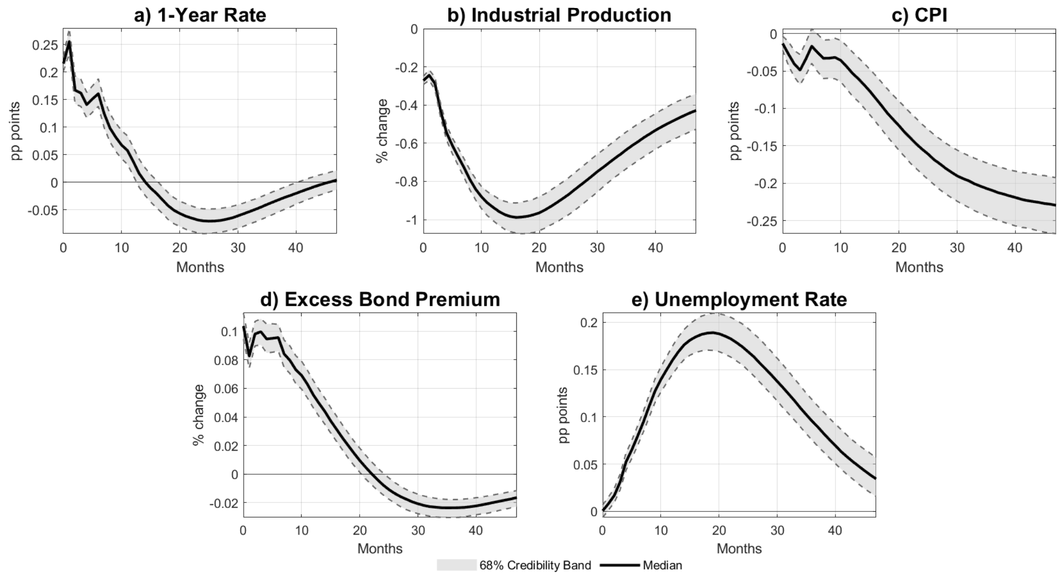

Figure A1: Estimated IRFs to a positive monetary policy shock: Proxy SVAR identification.

quency and estimate local projections of the quarterly BDS variables on these aggregated shocks. This approach avoids using a quarterly proxy directly in the Proxy SVAR, which may be problematic when the external instrument is originally observed at a higher frequency as aggregating the instrument prior to identification may weaken its relevance and blur the timing of the underlying policy shock. At the same time, it allows us to assess whether the responses of firm dynamics obtained in the main text are robust to an alternative, external-instrument-based identification of monetary policy shocks. The results from the local projections are reported in Figure A2.

C Value of the Marginal Worker

The value of the marginal worker in the JCC, given by equation (11), is defined as:

$$D(Z,z,n) = \tilde{\beta} E_{\lambda'|\lambda} \int_{\mathcal{S}} \left[\int_{\mathcal{Z}^{na}} J(Z',z',n) dH(z'|z) + \int_{\mathcal{Z}^{exp'}} \frac{\kappa}{q'} dH(z'|z) \right] dG. \tag{A1}$$

The integration over the survival region $\mathcal{S} = \{c^f : \tilde{F}(Z',z',n) > c^f\}$ ensures that the marginal value is weighted by the firm's survival probability.

The term inside the square brackets consists of two components: the marginal value of a worker in firms that do not adjust their labor force, and the corresponding value for workers

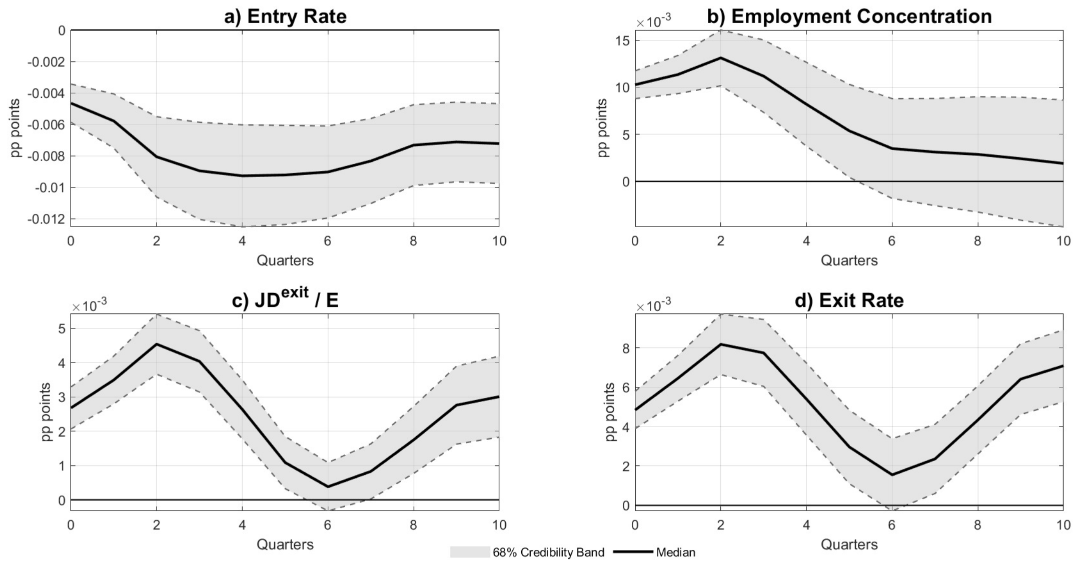

Figure A2: Local Projection on Proxy VAR shocks.

in expanding firms.

The regions of idiosyncratic shocks that determine the firm's employment adjustment policy are defined as follows:

$$\begin{cases} \mathcal{Z}^{f} & = \{z : z < \bar{z}_f\} \\ \mathcal{Z}^{na} & = \{z : \bar{z}_f \leq z < \bar{z}_{exp}\} \\ \mathcal{Z}^{exp} & = \{z : z \geq \bar{z}_{exp}\} \end{cases} \tag{A2}$$

where $\bar{z}_f = \max\{\mathbf{z}_{Z;n<n_{-1}}\}$ denotes the threshold of idiosyncratic productivity below which a firm with permanent productivity Z and initial size n finds it optimal to reduce its labor force. Conversely, $\bar{z}_{exp} = \min\{\mathbf{z}_{Z;n>n_{-1}}\}$ is the threshold above which the same firm expands its labor force.

Accordingly, $\mathcal{Z}^{na}$ is the region of shocks where the firm maintains its current employment level, while $\mathcal{Z}^{exp}$ is the region where the firm expands. By integrating the unconditional distribution $H(z'|z)$ over these restricted subsets, we obtain the submeasures required for equation (A1). This ensures that the terms are correctly weighted by the probability mass of the firm falling into each respective adjustment state.

As such, the different regions of the entire space $\mathcal{Z}$ in (A2) can be defined analogously as

$$\begin{cases} \mathcal{Z}^f & = \{z : n < n_{-1}\} \\ \mathcal{Z}^{na} & = \{z : n = n_{-1}\} \\ \mathcal{Z}^{exp} & = \{z : n > n_{-1}\} \end{cases} \tag{A3}$$

Therefore, the integrals in equation (A1) represent the expected future marginal value of current employment under two distinct scenarios. In the first case (non-adjustment), the marginal worker provides the future value J' (the marginal surplus). In the second case (expansion), the marginal worker provides a value equal to the saved marginal cost of hiring, κ/q', as the current worker reduces the need for future vacancy posting to reach the optimal firm size. Both outcomes are endogenously determined by the interaction of the firm's state and the realization of future idiosyncratic shocks.

Finally, note that in a model with a representative firm and no idiosyncratic shocks, equation (A1) collapses to $D = \kappa/q'$.

D Household's optimization problem

In this section, we derive the FOCs of the household's problem. There is one representative household, characterized by a measure $I = L + U$ of workers, either employed (L) or unemployed (U). The household supplies inelastically workers to the labor market as in Christiano et al. (2016). Each worker is risk-averse and receives the same amount of consumption by the household through perfect insurance. Lifetime utility is maximized with respect to consumption C, state-contingent asset B, and members of the family, $n(Z, z, n_{-1})$, working in a firm with idiosyncratic state denoted by the triplet (Z, z, n_{-1}).

The objective function of the representative household is

$$\mathbf{U}_t = E_0 \sum_{t=0}^{\infty} \beta^t u(C_t). \tag{A4}$$

Maximization is subject to the budget constraint:

$$C + B/P = \int_{\Lambda} wn \, d\Lambda + bU + \int_{\Lambda} \Pi \, d\Lambda - c_e N^e + B_{-1}R/P - T, \tag{A5}$$

where:

- $\int_{\Lambda} wn \, d\Lambda$ is total labor income: the firm-level wage bill wn integrated with respect to the measure Λ of operating firms (total mass $N = \int d\Lambda$);

- $\int_\Lambda \Pi \, d\Lambda$ are aggregate dividends, i.e. revenues minus labor, vacancies, and fixed costs summed across operating firms, gross of the total entry cost $c_e N^e$;
- b is unemployment benefit financed through government transfers T. Multiplying the total mass of unemployed workers U, it denotes total unemployment benefit expenditures;
- B represent investment in a nominal risk-free bond issued by the government;
- c_e is the entry cost in output units, multiplied by the mass of potential entrants N^e. Their product represents total investment in new firms by the representative household.

In setting its optimal choices of consumption and labor supply, the representative household takes into account the evolution of the stock of unemployed workers over time. The latter is affected by firm-level job destruction—induced by both idiosyncratic and aggregate shocks—along the intensive margin (separations in contracting firms) and the extensive margin (job destruction due to firm exit). Therefore, unemployment evolves according to the following law of motion:

$$U = \int_\Lambda \int_\mathcal{S} \int_{\mathcal{Z}^f} s n_{-1} dH(z|z_{-1}) dG d\Lambda_{-1} + \int_\Lambda \int_\mathcal{X} \int_\mathcal{Z} n_{-1} dH(z|z_{-1}) dG d\Lambda_{-1} + (1-\phi) U_{-1}, \quad \text{(A6)}$$

where the second term on the right-hand side measures the mass of firms exiting the market (i.e. which have a negative value net of fixed costs), and the integration over fixed costs $G(c^f)$ is partitioned into the survival region $\mathcal{S} = \{c^f : \tilde{F}(Z, z, n) > c^f\}$, and the exit region $\mathcal{X} = \{c^f : \tilde{F}(Z, z, n) \leq c^f\}$. Similarly, the integration over employment adjustments has been partitioned into a firing region $\mathcal{Z}^f = \{z : n < n_{-1}\}$, an expanding region $\mathcal{Z}^{exp} = \{z : n > n_{-1}\}$, and a non-adjustment region $\mathcal{Z}^{na} = \{z : n = n_{-1}\}$. The set $\mathcal{Z}$ represents the full support of the idiosyncratic shocks. Note that the second term on the right-hand side of equation (A6) measures the mass of workers from firms exiting the market; for these firms, the integration is performed over the entire distribution of idiosyncratic shocks[19]. By integrating the unconditional distributions G and H over these restricted subsets, the integrals represent submeasures that correctly weight the mass of firms in each state (survival/exit and adjustment/non-adjustment). Note that the variables must be integrated with respect to three distributions: the firm distribution at the end of the previous period Λ_{-1}; the distribution of fixed costs G, and the distribution of firm-level employment adjustments, H. Finally, $s = s(Z, z, n)$ is the endogenous, firm-specific, separation rate.

The representative household also has to account for the evolution of the mass of employed workers, which depends on each firm's idiosyncratic state. Fortunately, from the point of view of each worker, only three firm states matter: contracting (firing), non-adjusting, and expanding (hiring).

These states, in turn, determine the evolution of employed workers in the three situations:

1. the law of motion of workers employed in contracting firms:

$$n^f N^f = \int_{\Lambda} \int_{\mathcal{S}} \int_{\mathcal{Z}^{\{}} (1-s) n_{-1} dH(z|z_{-1}) dG d\Lambda_{-1}, \tag{A7}$$

   where the superscript f denotes firms that shed workers, and $\mathcal{Z}^{\{} = \{z : n < n_{-1}\}$ represents the region of idiosyncratic productivity shocks that lead a firm to reduce its labor force;

2. the law of motion of workers employed by firms that do not adjust their labor force:

$$n^{na} N^{na} = \int_{\Lambda} \int_{\mathcal{S}} \int_{\mathcal{Z}^{na}} n_{-1} dH(z|z_{-1}) dG d\Lambda_{-1}, \tag{A8}$$

   where the superscript na indicates firms that do not adjust their labor force, $\mathcal{Z}^{na} = \{z : n = n_{-1}\}$ represents the region of idiosyncratic productivity shocks that lead firms to freeze their employment adjustments;

3. the employment law of motion for workers employed in expanding firms[20]:

[19] Because the exit decision is triggered by the fixed cost realization c^f exceeding the firm's value, the firm ceases operations regardless of the specific shock $z \in \mathcal{Z}$ it would have otherwise faced.

[20] Notice that effective entrants always fall in this category.

$$n^{exp}N^{exp} = \int_{\Lambda}\int_{\mathcal{S}}\int_{\mathcal{Z}^{exp}} n_{-1}dH(z|z_{-1})dGd\Lambda_{-1} + \frac{qv^{exp}N^{exp}}{M}\phi U_{-1} \tag{A9}$$

where the superscript *exp* indicates firms that expand their labor force. $\mathcal{Z}^{exp} = \{z : n > n_{-1}\}$ represents the region of idiosyncratic productivity shocks that lead firm to expand their labor force. Finally, there are two additional terms in the equation that require further specification:

- $v^{exp}N^{exp} = \int_{\Lambda}\int_{\mathcal{S}}\int_{Z^{exp}} v(Z,z,n)dH(z|z_{-1})dGd\Lambda_{-1}$ and

$$\frac{qv^{exp}N^{exp}}{M} = \frac{q\int_{\Lambda}\int_{\mathcal{S}}\int_{\mathcal{Z}^{exp}} v(Z,z,n)dH(z|z_{-1})dGd\Lambda_{-1}}{M}$$

- $\frac{qv(Z,z,n)\,N(Z,z,n)}{M}$ denotes the fraction of total matches M attributed to firms characterized by the triplet (Z,z,n).[21] When multiplied by ϕU, it represents the mass of new hires in the (Z,z,n)-triplet class.

Finally, the representative household chooses the amount of labor to supply to firms, subject to the aggregate labor supply constraint:

$$\int_{\Lambda} nd\Lambda = L \tag{A10}$$

The first-order condition (FOC) for consumption is

$$\zeta = u(C) \tag{A11}$$

where ζ is the shadow value of consumption.

The FOC for the nominal risk-free bond B reads as:

$$\frac{1}{R} = \beta E_{\lambda'|\lambda}\frac{\zeta'}{\zeta}\frac{1}{\pi'}, \tag{A12}$$

leading to the usual Euler equation for consumption.

Given the preceding set of equations, the household chooses the number of workers to supply to each firm by maximizing (A4) with respect to n. For notational convenience, we

[21] In fact, $q_t v_{jt} = m_{jt}$ corresponds to the number of firm-specific matches.

suppress the dependence of variables on the triplet (Z, z, n), and denote $H^{(i)'} = H^{(i)}(z'|z)$. The corresponding FOC is:

$$W^{(i)} = \zeta w^{(i)} + E_{\lambda'|\lambda}\beta \Bigg\{ \int_{\mathcal{S}} \Bigg[\int_{\mathcal{Z}^f} (1 - s')W^{(f)'} dH'(z'|z) + \int_{\mathcal{Z}^{na}} W^{(na)'} dH'(z'|z) + \\ + \int_{\mathcal{Z}^{exp}} W^{(exp)'} dH'(z'|z) + \Gamma' \int s' dH'(z'|z) \Bigg] dG' + \Gamma' \int_{\mathcal{X}} \int_{\mathcal{Z}} dH'(z'|z) dG' \Bigg\}. \tag{A13}$$

Firm-level index $(i) = \{f, na, exp\}$ denotes the current employment adjustment policy at the individual firm—either firing, non-adjustment, or expanding—depending on the firm's state. Γ is the shadow value of non-work activities, whereas $\{W^{(f)}, W^{(n)a}, W^{(exp)}\}$ are the shadow values of workers employed in firing, non-adjusting or expanding firms. All of these values are firm-specific, and the wage w depends on each firm's idiosyncratic state (Z, z, n_{-1}).

In equation (A13) each worker is confronted with three distinct values of employment based on the current state and the anticipated wage $w^{(i)}$ in the future:

- $W^{(f)}$ - the value of employment when the firm finds it optimal to downsize, but the given worker is not dismissed;
- $W^{(na)}$ - the value of employment when the firm maintains the current number of workers;
- $W^{(exp)}$ - the value of employment when the firm hires new workers.

Also, the household chooses the allocation of its members to non-work activities denoted by U. The corresponding FOC is:

$$\Gamma = b\zeta + \beta E_{\lambda'|\lambda} \Bigg[(1 - \phi')\Gamma' + \int_{\Lambda} \int_{\mathcal{S}} \int_{\mathcal{Z}^{exp}} \frac{q'v'}{M'} \phi' W^{(exp)'} dH'(z'|z) dG' d\Lambda' \Bigg]. \tag{A14}$$

Equation (A14) represents the value of unemployment, incorporating both the expected value of being hired by a firm and the expected value of remaining unemployed, which each unemployed worker considers in their decision-making.

The household's marginal surplus from a match is equal to the difference between the shadow value of a worker employed at each firm $(W^{(i)})$ and the shadow value of unemployment (Γ), namely

$$H = W^{(i)} - \Gamma. \tag{A15}$$

E Wage Bargaining

The optimal wage schedule is determined through Nash bargaining between firms and workers[22]. Each firm and its workers maximize the joint surplus with respect to the wage w

$$\max_{w} J^{1-\eta} \left(\frac{H}{\zeta} \right)^{\eta} \tag{A16}$$

where the household's marginal value from a match with the individual firm is expressed in utils.

The first-order condition reads

$$\eta J = (1-\eta) \frac{H}{\zeta} \tag{A17}$$

which states that the marginal value of a match for workers and firms, weighted by their respective bargaining powers, must be equalized.

Substituting (A15) into (A17) yields

$$W^{(i)} - \Gamma = \frac{\eta}{(1-\eta)} J \zeta$$

indicating that the surplus of a worker being employed at any firm (left-hand side) depends on the value of the marginal worker in that firm.

Taking expectations of the previous equation and applying the JCC for hiring firms yields:

$$W^{(exp)'} - \Gamma' = \frac{\eta}{(1-\eta)} \frac{\kappa}{q'} \zeta' \tag{A18}$$

If instead the firm is separating from some of its workers

$$W^{(f)'} - \Gamma' = 0 \tag{A19}$$

[22] This Appendix follows the approach in Elsby and Michaels (2013)

Finally, in case of non-adjustments, the condition reads

$$W^{(na)'} - \Gamma' = \frac{\eta}{1-\eta} J'\zeta' \tag{A20}$$

Consider equation (A14) for Γ. By substituting (A18) into (A14), the value of unemployment in (A14) becomes

$$\Gamma = b\zeta + \beta E_{\lambda'|\lambda}\left[(1-\phi')\Gamma' + \int_{\Lambda}\int_{\mathcal{S}}\int_{\mathcal{Z}^{exp}} \frac{q'v'}{M'}\phi'\left(\Gamma' + \frac{\eta}{1-\eta}\frac{\kappa}{q'}\right) dH'dG'd\Lambda'\right]$$

Noting that $\int_{\Lambda} qvd\Lambda = M$, this can simplified as

$$\Gamma = \zeta b + \beta E_{\lambda'|\lambda}\Gamma' + \beta E_{\lambda'|\lambda}\frac{\eta}{1-\eta}\phi'\zeta'\frac{\kappa}{q'} \tag{A21}$$

so that the value of being unemployed is independent of firm heterogeneity and can be handled as in the representative-firm case. This result allows us to define the household's marginal surplus from a match, $H = W^{(i)} - \Gamma$, by subtracting (A21) from the corresponding value of employment (A13). The surplus then reads

$$\begin{aligned} H = \zeta(w^{(i)} - b) + \beta E_{\lambda'|\lambda}\Bigg\{ \int_{\mathcal{S}} \Bigg[\int_{\mathcal{Z}^f} (1-s')W^{(f)'}dH' + \int_{\mathcal{Z}^{na}} W^{(na)'}dH' + \\ + \int_{\mathcal{Z}^{exp}} W^{(exp)'}dH' + \Gamma'\int_{\mathcal{Z}^f} s'dH' \Bigg] dG' + \Gamma'\int_{\mathcal{X}}\int_{\mathcal{Z}} dH'dG' + \\ - \Gamma' - \frac{\eta}{1-\eta}\zeta'\phi'\frac{\kappa}{q'} \Bigg\} \end{aligned} \tag{A22}$$

Focusing on the expectation term inside the curly brackets, we can distinguish three main components. The first is the integral with respect to the cumulative distribution of fixed costs G in the survival region $\mathcal{S}$. This term is itself composed of four subcomponents, corresponding to the possible future states of an employed worker. The first three refer to continued employment at a shrinking firm, at a firm that freezes employment, or at an expanding firm. The fourth corresponds to the case in which the firm fires the worker, implying a transition from employment to unemployment. The remaining terms inside the curly brackets correspond to: (i) becoming unemployed due to firm exit, $\Gamma' \int_{\mathcal{X}} \int_{\mathcal{Z}} dH'dG'$, minus (ii) the continuation value of unemployed workers, $\Gamma' + \frac{\eta}{1-\eta}\zeta'\phi'\frac{\kappa}{q'}$.

Before proceeding with the derivation, note that

- $\int_{\mathcal{Z}^f} dH + \int_{\mathcal{Z}^{na}} dH + \int_{\mathcal{Z}^{exp}} dH = \int_{\mathcal{Z}} dH = 1$;
- $\int_{\mathcal{S}} dG + \int_{\mathcal{X}} dG = 1$;
- $W^{(f)} = \Gamma$ from (A19).

Therefore, by substituting (A18), (A19) (A20), and (A21) into (A22), H can be re-written as

$$
\begin{aligned}
H = \zeta(w^{(i)} - b) + \beta E_{\lambda'|\lambda}\Bigg\{ & \int_{\mathcal{S}} \Bigg[\int_{\mathcal{Z}^f} W^{(f)'} dH' + \int_{\mathcal{Z}^{na}} W^{(na)'} dH' + \\
& + \int_{\mathcal{Z}^{exp}} W^{(exp)'} dH' - \int_{\mathcal{Z}^f} s'(W^{(f)'} - \Gamma') dH' \Bigg] dG' + \\
& \Gamma' \int_{\mathcal{X}} \int_{\mathcal{Z}} dH' dG' + \\
& - \Gamma' \Bigg(\int_{\mathcal{X}} \int_{\mathcal{Z}} dH' dG' + \int_{\mathcal{S}} \int_{\mathcal{Z}} dH' dG' \Bigg) + \\
& - \frac{\eta}{1-\eta} \phi' \zeta' \frac{\kappa}{q'} \Bigg\}
\end{aligned}
$$

By manipulating this expression as follows

- $\Gamma' \int_{\mathcal{X}} \int_{\mathcal{Z}} dH' dG' - \Gamma' \int_{\mathcal{X}} \int_{\mathcal{Z}} dH' dG' = 0$
- $\int_{\mathcal{Z}^f} s'(W^{(f)'} - \Gamma') dH' dG' = 0$ since $W^{(f)} = \Gamma$
- $\int_{\mathcal{Z}} dH' = \int_{\mathcal{Z}^f} dH' + \int_{\mathcal{Z}^{na}} dH' + \int_{\mathcal{Z}^{exp}} dH'$

the result is

$$
\begin{aligned}
H = \zeta(w^{(i)} - b) + \beta E_{\lambda'|\lambda}\Bigg\{ \\
& \int_{\mathcal{S}}\left[\int_{\mathcal{Z}^f} W^{(f)'} dH' + \int_{\mathcal{Z}^{na}} W^{(na)'} dH' + \int_{\mathcal{Z}^{exp}} W^{(exp)'} dH'\right] dG' + \\
& - \Gamma' \int_{\mathcal{S}}\left[\int_{\mathcal{Z}^f} dH' + \int_{\mathcal{Z}^{na}} dH' + \int_{\mathcal{Z}^{exp}} dH'\right] dG' + \\
& - \frac{\eta}{1-\eta}\phi'\zeta'\frac{\kappa}{q'}\Bigg\}
\end{aligned}
$$

where all the terms referring to the values of being employed at different firm's state can be grouped with the value from unemployment Γ

$$
\begin{aligned}
H = & \zeta(w^{(i)} - b) + \beta E_{\lambda'|\lambda}\Bigg\{ \\
& \int_{\mathcal{S}}\left[\int_{\mathcal{Z}^f} (W^{(f)'} dH' - \Gamma') + \int_{\mathcal{Z}^{na}} (W^{(na)'} - \Gamma') dH' + \int_{\mathcal{Z}^{exp}} (W^{(exp)')} - \Gamma') dH'\right] dG' + \\
& - \frac{\eta}{1-\eta}\phi'\zeta'\frac{\kappa}{q'}\Bigg\}
\end{aligned}
$$

and finally, since $W^{(f)'} - \Gamma' = 0$ and $W^{(na)'} = H^{(na)'}$, $W^{(exp)'} = H^{(exp)'}$

$$
\begin{aligned}
H = & \zeta(w^{(i)} - b) + \beta E_{\lambda'|\lambda}\zeta'\Bigg\{ \\
& \frac{\eta}{1-\eta}\int\left[\int J' dH^{(na)'} + \int \frac{\kappa}{q'} dH^{(exp)'}\right] d\tilde{G}' + \\
& - \frac{\eta}{1-\eta}\phi'\frac{\kappa}{q'}\Bigg\}
\end{aligned}
$$

Hence, $\frac{H}{\zeta}$ - i.e. the household's marginal surplus from a match expressed in utils - can be expressed as

$$
H = w^{(i)} - b + E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\left\{\frac{\eta}{1-\eta}\int_{\mathcal{S}}\left[\int_{\mathcal{Z}^{na}} J' dH' + \int_{\mathcal{Z}^{exp}} \frac{\kappa}{q'} dH'\right] dG' - \frac{\eta}{1-\eta}\phi'\frac{\kappa}{q'}\right\} \tag{A23}
$$

where we substitute $E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}} = \beta\frac{u'(C')}{u'(C)} = \frac{\zeta'}{\zeta}$, i.e. the firms' real stochastic discount factor.

By combining the JCC given by

$$J = \alpha\rho Zzn^{\alpha-1} - \frac{\partial w^{(i)}}{\partial n}n - w^{(i)} + E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\int_{\mathcal{S}}\left[\int_{\mathcal{Z}^{na}} J'dH' + \int_{\mathcal{Z}^{exp}} \frac{\kappa}{q'}dH'\right]dG'$$

with equation (A23) in the first-order condition from wage bargaining $\frac{\eta}{(1-\eta)}J = \frac{H}{\zeta}$, we obtain

$$\frac{\eta}{1-\eta}\left[\alpha\rho Zzn^{\alpha-1} - \frac{\partial w^{(i)}}{\partial n}n - w^{(i)} + E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\int_{\mathcal{S}}\left(\int_{\mathcal{Z}^{na}} J'dH' + \int_{\mathcal{Z}^{exp}} \frac{\kappa}{q'}dH'\right)dG'\right] =$$
$$w^{(i)} - b + E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\left[\frac{\eta}{1-\eta}\int_{\mathcal{S}}\left(\frac{\kappa}{q'}\int_{\mathcal{Z}^{exp}} dH' + \int_{\mathcal{Z}^{na}} J'dH'\right)dG' - \phi'\frac{\kappa}{q'}\right]$$

that can be simplified by removing common terms, resulting into the following expression

$$\frac{\eta}{1-\eta}\left[\alpha\rho Zzn^{\alpha-1} - \frac{\partial w^{(i)}}{\partial n}n - w^{(i)} + \right] = w^{(i)} - b - E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\phi'\frac{\kappa}{q'}$$

Notice that we can suppress the index $(i) = \{f, na, exp\}$, since the wage depends on the firm-level employment level but not on the current employment adjustment policy at the firm - i.e. either firing, freezing, or hiring more workers.

This yields the following differential equation:

$$w = \eta\left[\alpha\rho Zzn^{\alpha-1} - \frac{\partial w}{\partial n}n + \kappa E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\theta'\right] + (1-\eta)b$$

which can be solved to obtain the wage equation in the text:

$$w = \eta\left[\frac{\alpha\rho Zzn^{\alpha-1}}{1-\eta(1-\alpha)} + \kappa E_{\lambda'|\lambda}\tilde{\boldsymbol{\beta}}\theta'\right] + (1-\eta)b. \tag{A24}$$

F Numerical solution

The steady state is solved numerically using a local approximation of the value function F_{jt}. The algorithm proceeds as follows:

- the idiosyncratic state $\{Z_j, z_{jt}, n_{jt-1}\}$ is discretized on a three-dimensional grid. Specifically, we set 4 nodes for Z_j, 19 for z_{jt}[23] and 80 for n_{jt-1};

[23]We use the Rouwenhorst's method to discretize AR(1) process.

- a first solution of the problem is obtained by iterating on the Bellman equation;
- if the solution does not hit the boundaries of the employment grid, the algorithm continues. Otherwise, the grid is enlarged, and the previous step is repeated;
- for each value of idiosyncratic productivity, we locally regress the objective function on the employment grid nodes immediately before and after the solution found in the previous step. The regressors are employment, employment squared, and a constant. Specifically:
  - let the objective function be $g_z(n)$. Consider the three nodes $[g_z(n_{k-1}), g_z(n_k), g_z(n_{k+1})]$, such that $n_{k-1} \leq n_k \leq n_{k+1}$, where n_k is the discrete maximizer of g_z on the grid $[n_{k-1}, n_k, n_{k+1}]$;
  - compute the OLS coefficients of

    $$g_z(n) = \beta_0 + \beta_1 n + \beta_2 n^2$$

    using the three nodes n_{k-1},n_k and n_{k+1} from the previous point;
  - take the first derivative of the interpolated g_z with respect to n and find the value of n that maximizes g_z by solving the first-order condition

    $$\tilde{n} = -\frac{\beta_1}{2\beta_2}$$

  - compute the corresponding objective function value: $\tilde{n}$:

    $$\tilde{g_z} = \beta_0 + \beta_1 \tilde{n} + \beta_2 \tilde{n}^2$$

- approximate F_j with $\tilde{g_z}$ computed for the firm with initial size equal to n_j;
- iterate until convergence of the value function.

To solve the dynamic stochastic equilibrium, we follow the method of Reiter (2009)[24]. Stochastic aggregate dynamics are computed by linearizing the system around the steady

[24] We use the MATLAB code developed by Costain and Nakov (2011).

state for each grid point. In this approach, the Bellman equation is treated not as a function but as a system of difference equations.

The model is represented by a system of non-linear equations:

$$E_t\mathbf{F}_t\Bigg(\mathbf{X}_t, \mathbf{X}_{t+1}, \log A_t, \log A_{t+1}\Bigg) = 0 \tag{A25}$$

where $\mathbf{X}_t$ comprises all model endogeneous variables, both firm-level and aggregate, and A_t represents an aggregate shock. At the steady state, equation (A25) reads

$$E_t\mathbf{F}\Bigg(\mathbf{X}, \mathbf{X}, 0, 0\Bigg) = 0 \tag{A26}$$

By computing the Jacobian of (A25) evaluated at the steady state equilibrium, the model can be approximated linearly for sufficiently small shocks:

$$E_t\mathcal{A}\big(\mathbf{X}_{t+1} - \mathbf{X}\big) + \mathcal{B}\big(\mathbf{X}_t - \mathbf{X}\big) + \mathcal{C}\log A_{t+1} + \mathcal{D}\log A_t \tag{A27}$$

where

- $\mathcal{A} = \frac{\partial \mathbf{F}_t}{\partial \mathbf{X}_{t+1}}$
- $\mathcal{B} = \frac{\partial \mathbf{F}_t}{\partial \mathbf{X}_t}$
- $\mathcal{C} = \frac{\partial \mathbf{F}_t}{\partial \log A_{t+1}}$
- $\mathcal{D} = \frac{\partial \mathbf{F}_t}{\partial \log A_t}$

This formulation allows the model to be solved using the QZ decomposition method of Klein (2000). In this way, we avoid imposing any bounded rationality assumptions, as required in the method of Krusell and Smith (1998).

G Penalty Function

Following the discussion in the main text, the limited information minimum distance objective is augmented with a penalty function $F^P(\cdot)$. This appendix provides its formal definition and interpretation.

For each firm size $s \in \{\text{new}, \text{small}, \text{large}\}$, define the vector of steady-state moments

$$m_s \equiv \begin{pmatrix} N^s/N^{Total} \\ L^s/L^{Total} \\ N^s_{exit}/N^s \end{pmatrix},$$

where N^s and L^s denote, respectively, the number of firms and total employment in firm class s. Let

$$\Delta m_s \equiv m_s^{Model} - m_s^{Data},$$

and stack deviations across firm sizes as

$$\Delta m \equiv \begin{pmatrix} \Delta m_{\text{new}} \\ \Delta m_{\text{small}} \\ \Delta m_{\text{large}} \end{pmatrix}.$$

The penalty function is defined as

$$\begin{aligned} F^P(\Delta M) = \|\Delta m\|_\infty &+ \left|s(\Omega^{DSGE}) - \bar{s}\right| \\ &+ \left|(N^{new}/N^{Total})^{Model} - (N^{new}/N^{Total})^{Data}\right| \\ &+ \left|(L^{small}/L^{Total})^{Model} - (L^{small}/L^{Total})^{Data}\right|. \end{aligned}$$

Each absolute value term measures the deviation of a model-implied steady-state moment from its empirical target. The first term applies an ℓ_∞ norm to the vector of size-specific moment deviations, penalizing the largest discrepancy across all margins of the stationary firm size distribution. This ensures that no single firm size or moment dimension exhibits an excessive mismatch between model and data.

The second term penalizes deviations of the aggregate separation rate. Here $s(\Omega^{DSGE})$ denotes the model-implied separation rate in the stationary equilibrium, and $\bar{s} = 0.09$ is its empirical target. This restriction is motivated by Elsby and Michaels (2013), who show that unemployment benefits interact with the dispersion of idiosyncratic shocks to shape equilibrium separation rates. Imposing this penalty ensures that the estimated model reproduces empirically plausible worker turnover in steady state.

The final two terms impose additional restrictions on the stationary distribution. The term

$$\left|(N^{new}/N^{Total})^{Model} - (N^{new}/N^{Total})^{Data}\right|$$

penalizes deviations between the model-implied and empirical firm entry rates, disciplining the relative mass of entrants with respect to incumbents. Without explicitly targeting this moment, the estimation may select parameter values that fit impulse responses while generating counterfactual entry rates, leading to distortions in the stationary firm size distribution.

Similarly, the term

$$\left|(L^{small}/L^{Total})^{Model} - (L^{small}/L^{Total})^{Data}\right|$$

penalizes deviations in the employment share of small firms. Given that firm shares across size classes are already disciplined by the moments included in Δm, targeting this moment anchors the allocation of employment across firm sizes and prevents the estimator from reallocating labor in ways that preserve firm counts while substantially distorting employment composition.

Overall, we verify that including these moments in the penalty function is sufficient to rule out parameter regions that generate large steady-state distortions, while leaving the fit of impulse responses essentially unchanged.

H Selection, Aggregate Supply, and Effective Productive Capacity: A General Example

This appendix provides a simple general framework for the mechanisms discussed in Section 8. The objective is to show that the response of prices and quantities to a demand contraction depends not only on how many firms exit, but also on which ones, thus, it depends on two endogenous supply-side margins. The first is a *shift effect*: entry and exit change effective productive capacity by changing the set of active firms, and thereby shifts aggregate supply inward. The second is a *slope effect*: conditional on the set of active firms, a change in aggregate output is redistributed across surviving incumbents, and this changes the slope of market supply through firms' marginal costs.

The key object is $H(P)$, which summarizes the contribution of active firms to aggregate supply, weighted by their productivity.

The appendix proceeds in two steps. It first presents a finite-firm economy, which makes the economic mechanism transparent. It then turns to a continuum of firms, where the same mechanism admits a derivative-based characterization.

H.1 Discrete economy

Consider a competitive industry with a finite number of firms indexed by $i = 1, \ldots, N$. Firm i has quadratic variable costs

$$C_i(q_i) = \frac{c_i}{2} q_i^2,$$

where a lower c_i denotes higher productivity. Marginal cost is therefore $MC_i(q_i) = c_i q_i$. A production function consistent with this cost structure is

$$q_i = \sqrt{\frac{2}{c_i}} \sqrt{n_i},$$

which implies

$$n_i = \frac{c_i}{2} q_i^2.$$

With the wage normalized to one, variable cost equals labor input, so labor demand and the quadratic cost function coincide.

Profit maximization yields the individual supply function

$$q_i(P) = \frac{P}{c_i}.$$

Substituting into labor demand gives firm-level employment

$$n_i(P) = \frac{c_i}{2} \left(\frac{P}{c_i} \right)^2 = \frac{P^2}{2c_i}.$$

Thus, both output and employment are weighted by the same term, $1/c_i$: more productive firms supply more and employ more.

Let $A(P)$ denote the set of firms active at price P. Aggregate supply is

$$Q(P) = \sum_{i \in A(P)} q_i(P) = P \sum_{i \in A(P)} \frac{1}{c_i}.$$

Define

$$H(P) \equiv \sum_{i \in A(P)} \frac{1}{c_i}.$$

Then aggregate supply can be written as

$$Q(P) = P\,H(P).$$

Aggregate employment is

$$N(P) = \sum_{i \in A(P)} n_i(P) = \frac{P^2}{2} \sum_{i \in A(P)} \frac{1}{c_i} = \frac{P^2}{2} H(P).$$

Equivalently,

$$H(P) = \frac{2N(P)}{P^2}.$$

This shows that $H(P)$ is naturally interpreted as an index of effective productive capacity: it rises when more productive firms are active and when activity is allocated toward them.

Demand is linear:

$$P = a - bQ.$$

Using $Q(P) = PH(P)$, equilibrium price satisfies

$$P = \frac{a}{1 + bH(P)}.$$

If the set of active firms is fixed, then $H(P)$ is fixed and

$$\frac{dP}{da} = \frac{1}{1 + bH(P)}.$$

The crucial issue is therefore how $H(P)$ changes when the shock changes the set of active firms.

H.1.1 Intensive-margin reallocation and the slope of market supply

The same object $H(P)$ also governs how a change in aggregate output is redistributed across surviving firms and, through that redistribution, the slope of market supply. Conditional on the active set $A(P)$, aggregate output can be written as $Q(P) = PH(P)$. Holding the active set fixed, this implies that the slope of inverse market supply is

$$\frac{dP}{dQ} = \frac{1}{H(P)}.$$

Hence, a larger $H(P)$ means that a given change in aggregate quantity requires a smaller change in the common price (equal to the common marginal cost), and therefore a flatter inverse supply schedule. This can also be seen at the firm level. Since $q_i(P) = P/c_i$, we have

$$q_i(P) = \frac{1/c_i}{H(P)} Q.$$

Therefore, conditional on the active set, any change in aggregate output is distributed across incumbents according to

$$\frac{dq_i}{dQ} = \frac{1/c_i}{H(P)}.$$

More productive firms absorb a larger share of the aggregate adjustment in levels because they account for a larger share of effective capacity. At the same time, any change in aggregate output operates through the movement in the common marginal cost. The cross-sectional allocation of the adjustment matters because it determines how much the common marginal cost must move for aggregate quantity to change.

This is the sense in which the distribution of production across surviving incumbents affects the slope of aggregate supply. Even for a given amount of entry or exit, two economies with different post-shock distributions of productive capacity can display different mappings from quantity adjustment into prices, because the same aggregate contraction can be spread over a different $H(P)$-weighted capacity base. In the main text, this is the second mechanism behind the heterogeneity counterfactual: not only which firms exit, but also how the residual contraction is absorbed by the firms that remain active.

H.1.2 Entry and exit in the discrete economy

Assume that each producing firm pays a fixed cost $F > 0$. At the optimum, profits are $\pi_i^* = \frac{P^2}{2c_i} - F$. A firm is active if and only if profits are non-negative, that is, if $\frac{P^2}{2c_i} \geq F$. Equivalently,

$$c_i \leq \bar{c}(P), \qquad \bar{c}(P) \equiv \frac{P^2}{2F}.$$

Thus, there is a unique profitability cutoff. Firms with sufficiently low cost, $c_i \leq \bar{c}(P)$, produce; firms with $c_i > \bar{c}(P)$ do not.

This same cutoff governs both exit and entry. For incumbents, a fall in P lowers $\bar{c}(P)$ and may force some firms with relatively high c_i to exit. For potential entrants, suppose each prospective entrant draws a value of c_i from a distribution $G(c)$. Entry occurs only if the draw satisfies $c_i \leq \bar{c}(P)$. Hence the probability of entry is $G(\bar{c}(P))$. A contractionary shock lowers P, lowers $\bar{c}(P)$, and therefore reduces the probability that a potential entrant draws a profitable productivity level. Entry does not create a lower cutoff. The active set is always characterized by a single upper cutoff in c-space: all firms with c_i below the threshold are active, and all firms with c_i above the threshold are inactive.[25]

In the finite-firm economy,

$$H(P) = \sum_{i=1}^{N} \frac{1}{c_i} \mathbf{1}\{c_i \leq \bar{c}(P)\}.$$

This is a step function of P. As P changes, $H(P)$ is constant as long as no threshold is crossed, and it jumps whenever a firm enters or exits. For this reason, $H'(P)$ is zero almost everywhere and undefined at the points where the active set changes. Therefore, in the discrete model the relevant object is not a local derivative but the finite change in effective productive capacity induced by a shock:

$$\Delta H = \sum_{i \in \text{entrants}} \frac{1}{c_i} - \sum_{i \in \text{exiters}} \frac{1}{c_i}.$$

A contractionary demand shock generates a larger price adjustment when it removes little H-weighted mass from the set of active firms, and a smaller price adjustment when it removes a large amount of H-weighted mass.

H.2 Continuum of firms

To characterize this mechanism with derivatives, it is useful to consider a continuum of firms. Let firms be indexed by their cost parameter $c \in (0, \infty)$, distributed with density $f(c)$ and

[25] In our dynamic model, the timing of entry matters. If entrants become active only in the following period, a change in the entry probability affects the future law of motion of active firms rather than contemporaneous $H(P)$. The static logic developed here abstracts from this timing distinction.

cumulative distribution function $F(c)$. The same profitability cutoff applies: $\bar{c}(P) = \frac{P^2}{2F}$. All firms with $c \leq \bar{c}(P)$ are active. In this case,

$$H(P) = \int_0^{\bar{c}(P)} \frac{1}{c} f(c)\, dc.$$

Aggregate supply is again $Q(P) = P\, H(P)$, aggregate employment is $N(P) = \frac{P^2}{2} H(P)$, and equilibrium price satisfies $a = P\big(1 + bH(P)\big)$. The probability that a potential entrant draws a profitable value of c is $G(\bar{c}(P))$. Thus, entry falls when P falls because the cutoff $\bar{c}(P)$ moves inward. If c is log-normal, say $\ln c \sim \mathcal{N}(\mu, \sigma^2)$, then the entry probability is

$$F(\bar{c}(P)) = \Phi\left(\frac{\ln \bar{c}(P) - \mu}{\sigma}\right),$$

which is increasing in P.

H.2.1 Derivative of $H(P)$ in the continuum

In the continuum case, Leibniz rule gives

$$H_P(P) = \frac{1}{\bar{c}(P)} f(\bar{c}(P))\, \bar{c}'(P).$$

Since $\bar{c}(P) = \frac{P^2}{2F}$, and $\bar{c}'(P) = \frac{P}{F}$, this simplifies to

$$H_P(P) = \frac{2}{P} f(\bar{c}(P)).$$

This expression makes the mechanism precise. The derivative of effective productive capacity depends on the density of firms at the profitability cutoff. For an infinitesimal change in price, only firms at the margin matter.

The corresponding derivative of the entry probability is

$$\frac{d}{dP} F(\bar{c}(P)) = f(\bar{c}(P))\, \bar{c}'(P) = \frac{P}{F} f(\bar{c}(P)).$$

Hence, a fall in P reduces both effective productive capacity and the probability of entry, and both effects are governed by the same upper cutoff.

H.2.2 Price response to a demand shock

Differentiating the equilibrium price $a = P\big(1 + bH(P)\big)$ with respect to a yields:

$$1 = \frac{dP}{da}\left[1 + bH(P) + bP\,H_P(P)\right].$$

Therefore,

$$\frac{dP}{da} = \frac{1}{1 + bH(P) + bP\,H_P(P)}.$$

Using $a/P = 1 + bH(P)$, the elasticity of price with respect to demand is

$$\frac{d\ln P}{d\ln a} = \frac{1 + bH(P)}{1 + bH(P) + bP\,H_P(P)}.$$

Substituting the expression for $H_P(P)$ gives

$$\frac{d\ln P}{d\ln a} = \frac{1 + bH(P)}{1 + bH(P) + 2b\,f(\bar{c}(P))}.$$

This shows that the attenuation of the price response is governed by the density of firms at the profitability cutoff. When many firms are near the margin, a small fall in demand reduces $H(P)$ more strongly, so less of the adjustment is absorbed by prices. When few firms are near the margin, $H(P)$ changes little, and prices absorb a larger share of the demand contraction. This explains the different response of prices shown in Panel e) of Figure 7

For finite shocks, the relevant object is not the local derivative but the total amount of H-weighted mass removed from activity. If price falls from P_0 to $P_1 < P_0$, the profitability cutoff declines from $\bar{c}(P_0)$ to $\bar{c}(P_1)$, and the corresponding fall in effective productive capacity is

$$H(P_0) - H(P_1) = \int_{\bar{c}(P_1)}^{\bar{c}(P_0)} \frac{1}{c} f(c)\,dc.$$

This makes clear that what matters is not heterogeneity per se, but how much H-weighted mass lies in the interval of firms pushed out of activity by the shock.

H.2.3 Interpretation

The discrete and continuous versions of the model deliver the same economic message. A contractionary demand shock lowers price. This reduces the profitability cutoff $\bar{c}(P)$. As a result, some incumbents exit and fewer potential entrants draw a sufficiently low c to enter.

Both margins reduce effective productive capacity. In addition, conditional on the set of firms that remains active, the contraction in aggregate output is redistributed across surviving incumbents in proportion to their contribution to effective capacity, and this determines how much the common marginal cost must move for the market to clear.

What matters for aggregate price adjustment is therefore not simply how dispersed firms are, but both (i) whether the firms affected by the movement in the cutoff account for a large or a small share of

$$H(P) = \sum \frac{1}{c_i} \qquad \text{or} \qquad H(P) = \int \frac{1}{c} f(c)\, dc,$$

and (ii) how the residual change in aggregate output is distributed across the surviving firms. If the firms that are pushed out by the shock contribute materially to effective productive capacity, then $H(P)$ falls strongly and the supply side contracts endogenously. If instead the marginal firms contribute little to effective productive capacity, then $H(P)$ changes little. Conditional on any given active set, the same object $H(P)$ also governs the slope of inverse market supply, because it determines how much the common marginal cost must move when surviving firms absorb the remaining adjustment in output.

This is the sense in which the two endogenous supply-side margins, i.e., shift and slope effects, jointly shape the aggregate supply response to demand shocks.

I Estimation of the Restricted Models used in the Counterfactual Analysis

For both restricted models used in the counterfactual—namely, the model featuring 1% employment concentration and the model with exogenous, size-independent exit—we re-estimate the relevant set of structural parameters using the same limited-information minimum distance procedure described in Section 5, targeting the same set of empirical impulse responses and applying the same penalty function to discipline the stationary distribution.

Table 5 reports the estimated parameters across the two specifications. The estimates are strikingly stable: all parameters remain virtually unchanged across models, with the most notable variation appearing in the Rotemberg price adjustment parameter ψ_R, which rises

Par.	Definition	Baseline	1% Emp. Conc.	Exog. Exit
κ	Cost of posting a vacancy	0.408	0.407	0.414
b	Unemployment benefit	1.303	1.303	1.301
λ_w	Degree of wage stickiness	0.975	0.976	0.978
σ_R	Std. deviation of MP shock	0.122	0.121	0.142
b_C	Strength of consumption habits	0.935	0.932	0.945
φ_R	Interest rate smoothing	0.731	0.731	0.731
φ_π	Policy rule parameter on inflation	1.513	1.511	1.508
φ_y	Policy rule parameter on output	0.013	0.013	0.014
ρ_R	Persistence of MP shock	0.854	0.855	0.852
ψ_R	Rotemberg price adjustment	102.0	104.4	115.6

Table 5: Estimated parameters across baseline and restricted model specifications.

Notes: All specifications are estimated using the limited-information minimum distance estimator described in Section 5, targeting the same set of empirical impulse responses. The restricted specifications impose either an alternative degree of employment concentration (1% of employment in large firms) or an exogenous, size-independent exit rate equal to 2.5% per quarter.

from 102.0 in the baseline to 104.4 under the 1% employment concentration specification and 115.6 under the exogenous exit specification. When there are few small, fragile firms, the model requires a higher degree of price rigidity to match the empirical impulse responses.

The standard deviation of the monetary policy shock, σ_R, also increases modestly under the exogenous exit specification, from 0.122 to 0.142, consistent with the interpretation that a larger shock is required to fit the same empirical responses once the amplification mechanism through endogenous exit is removed.

This parameter stability has an important implication for the counterfactual design: differences in the impulse responses documented in Sections 8.2 and 8.1 are largely attributable to the imposed structural restrictions—differences in the degree of employment concentration or in the nature of the exit margin—rather than to changes in estimated parameters. This provides a clean identification of the role of each mechanism in shaping the transmission of monetary policy. To illustrate this, consider figures A3 and A4 which represent the counterparts of Figures 7 and 6, respectively, where parameters of the restricted models take their estimated values. The difference in the estimated value of the Rotemberg parameter leaves the analysis in the main text unchanged.

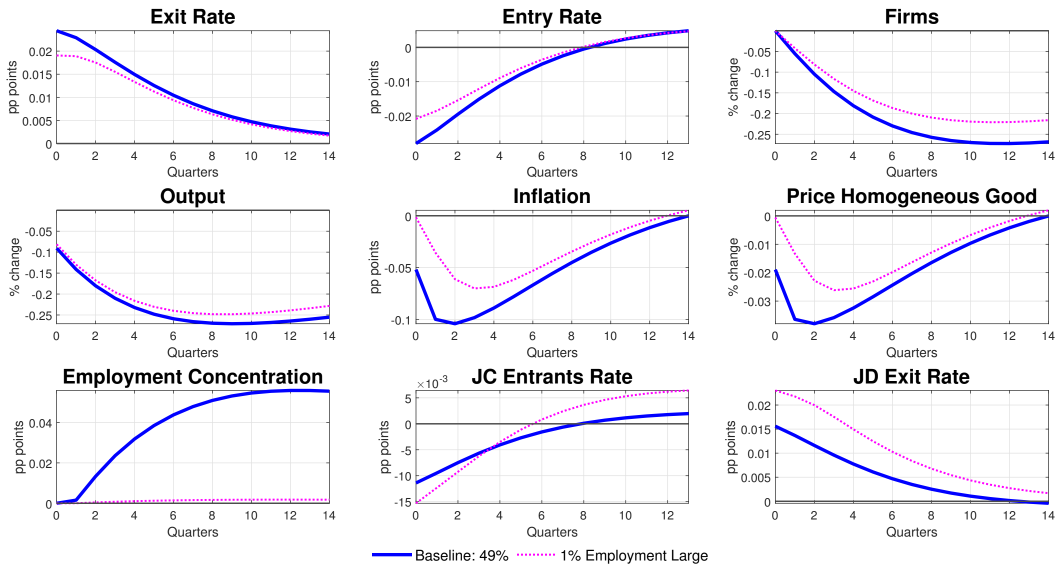

Figure A3: IRFs under alternative degrees of employment concentration with re-estimated parameters).

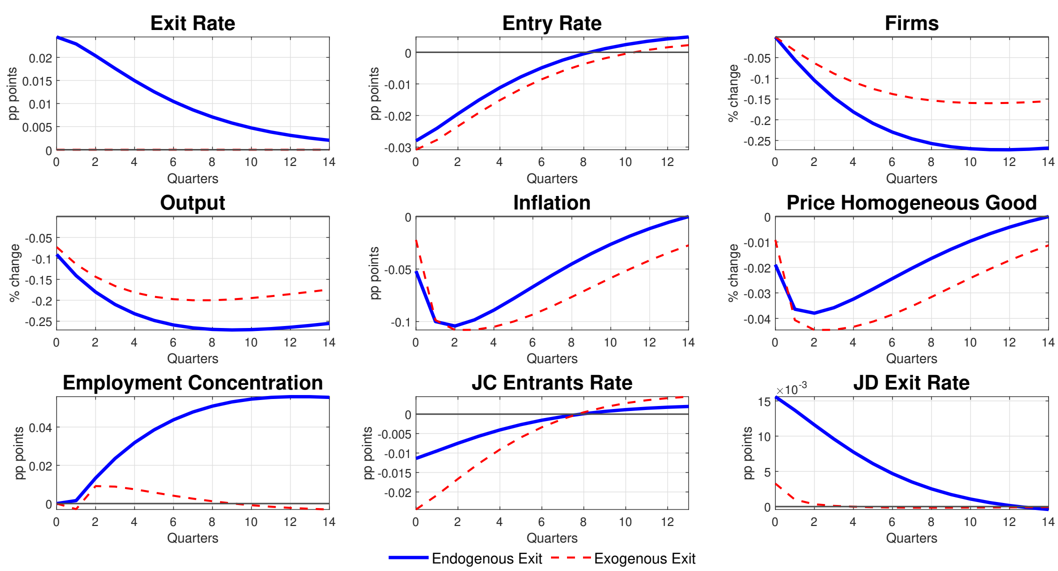

Figure A4: IRFs under exogenous and endogenous exit with re-estimated parameters

References

Bahaj, S., Foulis, A. and Pinter, G.: 2022, The asymmetric effects of monetary policy on firm employment, *Journal of Monetary Economics* .

Barnichon, R. and Mesters, G.: 2021, The phillips multiplier, *Journal of Monetary Economics* **117**, 689–705.

Bergin, P. R. and Corsetti, G.: 2008, The extensive margin and monetary policy, *Journal of Monetary Economics* **55**(7), 1222–1237.

Bergin, P. R., Feng, L. and Lin, C.-Y.: 2018, Firm entry and financial shocks, *The Economic Journal* **128**(609), 510–540.

Bilal, A., Engbom, N., Mongey, S. and Violante, G. L.: 2022, Firm and worker dynamics in a frictional labor market, *Econometrica* **90**(4), 1425–1462.

Bilbiie, F. O.: 2020, Monetary neutrality with sticky prices and free entry, *Review of Economics and Statistics* pp. 1–42.

Bilbiie, F. O., Ghironi, F. and Melitz, M. J.: 2012, Endogenous entry, product variety, and business cycles, *Journal of Political Economy* **120**(2), 304–345.

Bilbiie, F. O., Ghironi, F., Melitz, M. J., Midrigan, V. and Rotemberg, J. J.: 2007, Monetary policy and business cycles with endogenous entry and product variety [with comments and discussion], *NBER Macroeconomics Annual* **22**, 299–379.

Blanchard, O. and Galí, J.: 2010a, Labor markets and monetary policy: A new keynesian model with unemployment, *Americal Economic Journal: Macroeconomics* **2**(1-30).

Blanchard, O. and Galí, J.: 2010b, Labor markets and monetary policy: A new keynesian model with unemployment, *American economic journal: macroeconomics* **2**(2), 1–30.

Boivin, J. and Giannoni, M. P.: 2006, Has monetary policy become more effective?, *The Review of Economics and Statistics* **88**(3), 445–462.

Canova, F. and Ferroni, F.: 2020, A hitchhiker's guide to empirical macro models. mimeo.

Christiano, L. J., Eichenbaum, M. and Evans, C. L.: 2005, Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of political Economy* **113**(1), 1–45.

Clementi, G. L. and Palazzo, B.: 2016, Entry, exit, firm dynamics, and aggregate fluctuations, *American Economic Journal: Macroeconomics* **8**(3), 1–41.

Colciago, A. and Etro, F.: 2010, Endogenous market structures and the business cycle, *The Economic Journal* **120**(548), 1201–1233.

Colciago, A., Fasani, S. and Rossi, L.: 2025, Firm entry, endogenous wage moderation, and labor market dynamics, *European Economic Review* **172**, 104939.

Colciago, A., Lindenthal, V. and Trigari, A.: 2019, Who creates and destroys jobs over the business cycle?

Colciago, A. and Silvestrini, R.: 2022, Monetary policy, firm heterogeneity, and market concentration, *European Economic Review* **144**, 104089.

Costain, J. and Nakov, A.: 2011, Price adjustments in a general model of state-dependent pricing, *Journal of Money, Credit and Banking* **43**(2-3), 385–406.

Davis, S. J. and Haltiwanger, J.: 1992, Gross job creation, gross job destruction, and employment reallocation, *The Quarterly Journal of Economics* **107**(3), 819–863.

Davis, S. J., Haltiwanger, J. C. and Schuh, S.: 1996, Job creation and destruction mit press, *Cambridge, MA* .

Debortoli, D., Forni, M., Gambetti, L. and Sala, L.: 2023, Asymmetric monetary policy tradeoffs, *CEPR Discussion Paper No. 18438* .

Elsby, M. W. L. and Michaels, R.: 2013, Marginal jobs, heterogeneous firms, and unemployment flows, *American Economic Journal: Macroeconomics* **5**(1), 1–48.

Flinn, C. J.: 2006, Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates, *Econometrica* **74**(4), 1013–1062.

Fort, T. C., Haltiwanger, J., Jarmin, R. S. and Miranda, J.: 2013, How firms respond to business cycles: The role of firm age and firm size, *IMF Economic Review* **61**(3), 520–559.

Gabriel, R. D.: 2023, Monetary policy and the wage inflation-unemployment tradeoff, *European Economic Review* **159**, 104565.

Gertler, M. and Gilchrist, S.: 1994, Monetary policy, business cycles, and the behavior of small manufacturing firms, *Quarterly Journal of Economics* **109**(2), 309–340.

Gertler, M. and Karadi, P.: 2015, Monetary policy surprises, credit costs, and economic activity, *American Economic Journal: Macroeconomics* **7**(1), 44–76.

Gertler, M. and Trigari, A.: 2009, Unemployment fluctuations with staggered nash wage bargaining, *Journal of Political Economy* **117**(1), 38–86.

Ghironi, F. and Melitz, M. J.: 2005, International trade and macroeconomic dynamics with heterogeneous firms, *The Quarterly Journal of Economics* **120**(3), 865–915.

Gutiérrez, G., Jones, C. and Philippon, T.: 2021, Entry costs and aggregate dynamics, *Journal of Monetary Economics* **124**, S77–S91.

Hagedorn, M. and Manovskii, I.: 2008, The cyclical behavior of equilibrium unemployment and vacancies revisited, *American Economic Review* **98**(4), 1692–1706.

Hall, R. E.: 2005, Employment fluctuations with equilibrium wage stickiness, *American Economic Review* **95**(1), 50–65.

Haltiwanger, J., Jarmin, R. S. and Miranda, J.: 2013, Who creates jobs? small versus large versus young, *Review of Economics and Statistics* **95**(2), 347–361.

Hamano, M. and Zanetti, F.: 2017, Endogenous product turnover and macroeconomic dynamics, *Review of Economic Dynamics* **26**, 263–279.

Hopenhayn, H. A.: 1992, Entry, Exit, and Firm Dynamics in Long Run Equilibrium, *Econometrica* **60**(5), 1127–1150.
URL: *https://ideas.repec.org/a/ecm/emetrp/v60y1992i5p1127-50.html*

Jaef, R. N. F. and Lopez, J. I.: 2014, Entry, trade costs, and international business cycles, *Journal of International Economics* **94**(2), 224–238.

Jaimovich, N. and Floetotto, M.: 2008, Firm dynamics, markups, and aggregate productivity, *Journal of Monetary Economics* **55**(3), 408–427.

Jarociński, M. and Karadi, P.: 2020, Deconstructing monetary policy surprises - the role of information shocks, *American Economic Journal: Macroeconomics* **12**(2), 1–43.

Jordà, Ò.: 2005, Estimation and inference of impulse responses by local projections, *American economic review* **95**(1), 161–182.

Lewis, V. and Poilly, C.: 2012, Firm entry, markups and the monetary transmission mechanism, *Journal of Monetary Economics* **59**(7), 670–685.

Melitz, M. J.: 2003, The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* **71**, 1695–1725.

Miranda-Agrippino, S.: 2016, Unsurprising shocks: Information, premia, and the monetary transmission, *Bank of England, Working Paper n. 626* .

Miranda-Agrippino, S. and Ricco, G.: 2021, The transmission of monetary policy shocks, *American Economic Journal: Macroeconomics* **13**(3), 74–107.

Mortensen, D. T. and Pissarides, C. A.: 1994, Job creation and job destruction in the theory of unemployment, *The Review of Economic Studies* **61**(3), 397–415.
URL: *http://dx.doi.org/10.2307/2297896*

Moscarini, G. and Postel-Vinay, F.: 2012, The contribution of large and small employers to job creation in times of high and low unemployment, *American Economic Review* **102**(6), 2509–2539.

Ottonello, P. and Winberry, T.: 2020, Financial heterogeneity and the investment channel of monetary policy, *Econometrica* **88**(6), 2473–2502.

Popov, A. and Steininger, M.: 2023, Monetary policy and market competition: Evidence from the euro area, *European Economic Review* .

Reiter, M.: 2009, Solving heterogeneous-agent models by projection and perturbation, *Journal of Economic Dynamics and Control* **33**(3), 649–665.

Rossi, L.: 2019, The overshooting of firms' destruction, banks and productivity shocks, *European Economic Review* **113**, 136–155.

Rotemberg, J.: 1997, An optimization-based econometric framework for the evaluation of monetary policy, *NBER Macroeconomics Annual/The MIT Press* .

Rotemberg, J. J.: 1982, Sticky prices in the united states, *Journal of political economy* **90**(6), 1187–1211.

Sedláček, P.: 2020, Lost generations of firms and aggregate labor market dynamics, *Journal of Monetary Economics* **111**, 16–31.

Shimer, R.: 2005, The cyclical behavior of equilibrium unemployment and vacancies, *American Economic Review* **95**(1), 25–49.

Siemer, M.: 2014, Financial shocks, firm entry, and business cycles, *Journal of Economic Dynamics and Control* .

Singh, P., Suda, K. and Zervou, F.: 2022, Firm size distribution and asymmetric effects of monetary policy, *Journal of Economic Dynamics and Control* .

Sterk, V., Sedláček, P. and Pugsley, B.: 2021, The nature of firm growth, *American Economic Review* **111**(2), 547–79.

Stole, L. A. and Zwiebel, J.: 1996, Intra-firm bargaining under non-binding contracts, *The Review of Economic Studies* **63**(3), 375–410.

Trigari, A.: 2009, Equilibrium unemployment, job flows, and inflation dynamics, *Journal of money, credit and banking* **41**(1), 1–33.