

DISCUSSION PAPER SERIES

DP15500
(v. 3)

A Congestion Theory of Unemployment Fluctuations

Yusuf Mercan, Benjamin Schoefer and Petr Sedláček

MONETARY ECONOMICS AND FLUCTUATIONS

CEPR

A Congestion Theory of Unemployment Fluctuations

Yusuf Mercan, Benjamin Schoefer and Petr Sedláček

Discussion Paper DP15500
First Published 27 November 2020
This Revision 13 March 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Yusuf Mercan, Benjamin Schoefer and Petr Sedláček

A Congestion Theory of Unemployment Fluctuations

Abstract

We propose a theory of unemployment fluctuations in which new hires and incumbent workers are imperfect substitutes. Hence, attempts to hire away the unemployed during recessions diminish the marginal product of new hires, discouraging job creation. This single feature achieves a ten-fold increase in the volatility of hiring in an otherwise standard search model, produces a realistic Beveridge curve despite countercyclical separations, and explains 30-40% of US unemployment fluctuations. Additionally, it explains the excess procyclicality of new hires' wages, the cyclical labor wedge, countercyclical earnings losses from job displacement, and the limited steady-state effects of unemployment insurance.

JEL Classification: J64, J63, E24

Keywords: N/A

Yusuf Mercan - yusuf.mercan@unimelb.edu.au
University of Melbourne

Benjamin Schoefer - schoefer@berkeley.edu
University California, Berkeley and CEPR

Petr Sedláček - petr.sedlacek@economics.ox.ac.uk
University of New South Wales and CEPR

A Congestion Theory of Unemployment Fluctuations*

Yusuf Mercan
University of Melbourne

Benjamin Schoefer
UC Berkeley

Petr Sedláček
UNSW & CEPR

March 2022

Abstract

We propose a theory of unemployment fluctuations in which new hires and incumbent workers are imperfect substitutes. Hence, attempts to hire away the unemployed during recessions diminish the marginal product of new hires, discouraging job creation. This single feature achieves a ten-fold increase in the volatility of hiring in an otherwise standard search model, produces a realistic Beveridge curve despite countercyclical separations, and explains 30–40% of US unemployment fluctuations. Additionally, it explains the excess procyclicality of new hires' wages, the cyclical labor wedge, countercyclical earnings losses from job displacement, and the limited steady-state effects of unemployment insurance.

*We thank the editor Ayşegül Şahin, and three anonymous reviewers, as well as Chris Edmond, Michael Elsby, Domenico Ferraro, Shigeru Fujita, Gabriel Chodorow-Reich, Bart Hobijn, Philip Jung, Fatih Karahan, Guido Menzio, Simon Mongey, Emi Nakamura, Moritz Schularick, Robert Shimer, and Jón Steinsson for useful comments. We also thank seminar audiences at Aarhus University, Arizona State University, Boston University, Columbia University, Copenhagen-Macro, Deakin University, Erasmus Universiteit Rotterdam, Federal Reserve Bank of Philadelphia, Federal Reserve Bank of New York, Federal Reserve Bank of Richmond, Harvard University, Hitotsubashi University, LAEF Conference on Labor Markets and Macroeconomic Outcomes, 11th ifo Conference on Macroeconomics and Survey Data, Massachusetts Institute of Technology, New York University, New York University Abu Dhabi, Reserve Bank of Australia, UC Berkeley, University of Melbourne, University of Adelaide and Waseda University. Sedláček gratefully acknowledges financial support from the European Research Council of the European Commission [grant number 802145]. Earlier stages of this project were also financially supported by the German Research Foundation (DFG) in 2016-2017 [grant number: SE 2554/1-1]. Nicholas Sander provided excellent research assistance. Mercan: yusuf.mercan@unimelb.edu.au, Schoefer: schoefer@berkeley.edu, Sedláček: p.sedlacek@unsw.edu.au.

1 Introduction

Recessions are times when labor demand plummets and unemployment increases. However, the dominant framework for studying equilibrium unemployment, the Diamond-Mortensen-Pissarides (DMP) search and matching model, has a hard time generating realistically deep recessions following negative shocks to labor demand.¹ A main challenge for generating realistic labor market volatility in this model is inherent to the search frictions that are at its core: the abundance of unemployed job seekers lowers the search costs of hiring, which incentivizes job creation, attenuating the ultimate increase in unemployment.

We propose a parsimonious yet powerful refinement of the DMP model that curbs this attenuation mechanism: a constant returns to scale aggregate production function that exhibits diminishing returns to new hires, a feature we call *congestion in hiring*. Specifically, new hires out of unemployment are imperfect substitutes for incumbent workers, and more generally workers are heterogeneous production factors distinguished by their hiring cohort. This modeling choice is meant to stand in for various potential micro foundations of congestion, while nesting the standard DMP model of Shimer (2005) as a special case.² In that standard model, all workers are perfectly substitutable and, therefore, firms can absorb the unemployed in recessions without compromising on their productivity.

Our congestion mechanism paints a new picture of recessions. Like in the standard model, when unemployment increases, as hiring falls and job separations rise, firms incipiently do take advantage of cheaper hiring by absorbing some of the additional unemployed. As a result, the gross flows from unemployment back into employment increase, and recently unemployed workers become abundant in the workforce.³ Unlike in the standard model, however, exactly this increase in hiring diminishes the marginal product of labor in the jobs the unemployed fill, which discourages further job creation. That is, rather than asking why firms hire so little, our paper points out that firms have already absorbed so many of the unemployed that the jobs they would fill are crowded.

This congestion mechanism improves the business cycle performance of the DMP model considerably. It raises the volatility of labor market tightness ten-fold, to empirically realistic levels. It produces a realistic Beveridge curve despite countercyclical separations. On its own, it accounts for around 30–40% of US unemployment fluctuations and much of its persistence. In addition, the model accounts for a range of other business cycle patterns linked to unemployment: the excess

¹See, e.g., Shimer (2005); Hall (2005b); Hagedorn and Manovskii (2008); Gertler and Trigari (2009); Pissarides (2009); Christiano, Eichenbaum, and Trabandt (2016); Hall (2017); Ljungqvist and Sargent (2017); Christiano, Eichenbaum, and Trabandt (2020).

²For example, different hiring cohorts may be on different rungs of the career ladder, have different skills, and perform different tasks; other sources of such congestion include internal labor markets, in which entry level-jobs (“ports of entry”) are imperfect substitutes to higher-tier positions (see, e.g., Doeringer and Piore, 1985; Lazear and Oyer, 2004), or human capital accumulation while on the job and skill loss in unemployment (see, e.g., Ljungqvist and Sargent, 1998; Kroft, Lange, and Notowidigdo, 2013; Kehoe, Lopez, Midrigan, and Pastorino, 2019).

³Such countercyclical unemployment-to-employment (UE) flows are an empirical fact in the US and other OECD countries (see, e.g., Blanchard and Diamond, 1990; Burda and Wyplosz, 1994; Fujita and Ramey, 2009; Elsby, Hobijn, and Şahin, 2013). To rationalize countercyclical UE flows, our model features countercyclical separation rates (as in the data), whereas the standard assumption of constant separation rates would imply counterfactually procyclical UE flows.

procyclicality of wages of newly hired workers compared to average wages, the countercyclical labor wedge, large countercyclical earnings losses from displacement and from labor market entry, and the long-run insensitivity of unemployment to policies such as unemployment insurance.

We begin our exposition with a simple two-equation DMP model as a stepping stone to our full quantitative analysis. Namely, we introduce a distinction between new hires and incumbent workers, who are imperfect substitutes in production. We then derive a key prediction of congestion in hiring: the elasticity of job creation with respect to a sudden increase in unemployment—brought about by, e.g., a spike in separations—is negative only with imperfectly substitutable workers. Otherwise, under perfect substitution, a rise in unemployment is unequivocally good news for firms: it lowers the search costs of hiring and spurs job creation—the typical pitfall of the standard DMP framework, in which separation shocks lead to expansions in vacancy creation, leading to a counterfactually positively sloped Beveridge curve (e.g., Shimer, 2005). In our model, however, increased unemployment-to-employment (UE) flows push down the marginal product of new hires, leading vacancies to respond less than one-for-one with unemployment.

To substantiate our congestion mechanism, we present empirical evidence that labor market tightness (the ratio of vacancies to unemployed job seekers) falls in response to separation rate shocks. Namely, we use a vector autoregressive (VAR) model to estimate the impulse response function (IRF) of labor market tightness to unexpected increases in unemployment brought about by innovations to the separation rate. This empirical response is strongly negative in the US data, consistent with our congestion model but in stark contrast to the zero response predicted by the standard DMP model. Given the stark difference between the prediction of the standard theory and data, we use the empirical IRF as a target to discipline the key parameter guiding the degree of congestion in our full model.⁴ This parameter determines the elasticity of substitution between cohorts in a constant elasticity of substitution production function that exhibits overall constant returns to scale. Additionally, we externally validate our estimated congestion parameter value with empirical evidence on the more pronounced procyclicality of wages of new hires relative to incumbent workers in the US data (see, e.g. Pissarides, 2009), which our model’s congestion mechanism rationalizes with its amplification of new hires’ productivity.

In contrast to the standard DMP model (Shimer, 2005), our full model with countercyclical congestion and standard total factor productivity and separation shocks, replicates key business cycle patterns of labor market variables successfully. For example, the standard deviation of labor market tightness is 90% of that in the data, and the correlation between unemployment and vacancies, i.e. the slope of the Beveridge curve, is -0.716 in our model compared to -0.934 in the data. Furthermore, countercyclical congestion alone accounts for 30–40% of US unemployment fluctuations, and accounts for most of its persistence. When we surgically turn off congestion (by setting worker types to be perfect substitutes), our calibrated model precisely nests the standard DMP model and exhibits its well-known cyclical failures (Shimer, 2005): counterfactually smooth

⁴Here, our calibration strategy echoes the important work by Coles and Moghaddasi Kelishomi (2018) (Table 4), who propose a DMP model modifying free entry in vacancies.

business cycles and a wrongly (positively) sloped Beveridge curve.

The quantitative success of our model rests on two key aspects of our congestion mechanism. First, the endogenous productivity of new hires is considerably (roughly five times) more volatile than average labor productivity. This is because when productivity is low, UE flows are typically high, lowering the marginal product of new hires even further. Second, cohort effects make aggregate conditions at the time of hiring have long-lasting effects on new hires' productivity. Conversely, the model's performance does *not* rest on choices such as wage rigidity (see, e.g., Shimer, 2004; Hall, 2005b; Michailat, 2012; Schoefer, 2021) or a small fundamental surplus (Hagedorn and Manovskii, 2008; Ljungqvist and Sargent, 2017).⁵

In addition, by offering a new perspective on unemployment fluctuations, our framework proposes a unified solution to three other, related long-standing macroeconomic issues. These resolutions provide further external validity to our productivity-based modeling of congestion. First, countercyclical congestion—in particular the cyclical difference between average productivity and that of new hires—provides a quantitative explanation for the countercyclical labor wedge, i.e., the gap between the marginal rate of substitution between consumption and leisure, and the marginal product of labor implied by a standard Real Business Cycle (RBC) model (Chari, Kehoe, and McGrattan, 2007; Shimer, 2009; Karabarbounis, 2014; Bils, Klenow, and Malin, 2018). Second, our model features large, countercyclical, and persistent earnings losses from job displacement (Davis and von Wachter, 2011) and from labor market entry such as from university graduation (Kahn, 2014; Oreopoulos, von Wachter, and Heisz, 2012; Schwandt and von Wachter, 2019)—consistent with the cohort-specific productivity channel. Third, our model generates amplification through more volatile allocative productivity, rather than raising the long-run elasticity of labor market tightness to productivity. Hence, we overcome the critique raised by Costain and Reiter (2008), that standard DMP models cannot *simultaneously* exhibit realistic, productivity-driven, business cycles and a low sensitivity to unemployment insurance (UI) generosity.

Our paper relates to recent work on the gradual reduction in unemployment during recoveries (Dupraz, Nakamura, and Steinsson, 2019; Hall and Kudlyak, 2020b,a). A notion of congestion is also present in important prior work by Coles and Moghaddasi Kelishomi (2018), who modify the assumption of free entry in vacancy creation in a standard DMP model and highlight the role of separation shocks in labor market fluctuations. Hall (2005a) and Engbom (2020) propose models in which the unemployed send applications less selectively in recessions, such that recruitment becomes more costly, a process that can be interpreted to reflect congestion. Michailat (2012) presents a model with wage rigidity and diminishing returns in *total* employment. That model does not exhibit congestion in hiring (i.e., it would predict essentially no effect of separation rate shocks

⁵In fact, even models with structurally different congestion mechanisms generate the same degree of amplification as our baseline model as long as they are calibrated to match the empirical degree of congestion (the decline in labor market tightness in response to separation shocks as evidenced by our VAR estimation). We establish robustness regarding labor market amplification to these “iso-congestion” models using the example of convex adjustment costs (Fujita and Ramey, 2007; Coles and Moghaddasi Kelishomi, 2018; Mercan and Schoefer, 2020), and by allowing only a subset of new hires to generate congestion reminiscent of models of turbulence *without congestion* (e.g., Ljungqvist and Sargent, 1998, 2004; den Haan, Haefke, and Ramey, 2005).

on labor market tightness), although it rationalizes a rat-race effect in net employment (Landais, Michaillat, and Saez, 2018). Our model also speaks to the effects of reallocation shocks and churn more generally (see, e.g., Lilien, 1982; Abraham and Katz, 1986; Chodorow-Reich and Wieland, 2020), as congestion slows down firms’ absorption of separated workers. Finally, Eyigungor (2010) studies a DMP model with vintage effects in capital and embodied technology, thereby separating productivity in new jobs from that in old ones.

The paper is structured as follows. In Section 2, we introduce our notion of congestion in hiring via a simple model and interpret empirical evidence pointing to countercyclical UE flows and congestion through the lens of this framework. Section 3 then presents our full quantitative model. Section 4 discusses calibration, and Section 5 investigates the model’s business cycle performance. Section 6 studies three further macro implications of congestion. Section 7 concludes.

2 The Mechanism and Empirical Evidence

In this section, we introduce our mechanism using an extension to the simplest, two-equation, DMP model. Within this stylized model, we derive analytical expressions which help define our congestion channel and provide intuition for its sources. In addition, we present empirical evidence consistent with our analytical predictions and develop a set of moments which we use to discipline our full quantitative model in the subsequent sections.

2.1 A Simple DMP Model with Congestion in Hiring

We begin by revisiting the standard DMP model, which is characterized by two equilibrium conditions determining unemployment u and job creation (vacancies) v .

A Two-equation DMP Model. The first condition is based on worker flows and pins down unemployment. Using subscripts $+1$ and -1 to denote leads and lags of variables, the law of motion for unemployment, fixing the labor force to one, is given by

$$u_{+1} = (1 - f(\theta)) u + \delta(1 - u), \tag{1}$$

where δ is the job separation rate at which workers flow into unemployment, and $f(\theta)$ is the job finding rate at which unemployed job searchers find and accept a job. The latter is determined by labor market tightness θ , the ratio of vacancies v to unemployment u . The implied steady-state unemployment rate—given by equating flows into and out of unemployment—is $u = \frac{\delta}{\delta + f(\theta)}$.

The second condition—the canonical free-entry condition—pins down job creation and, in turn, the job finding rate. Job creation is such that expected costs and benefits of hiring are equal

in present value:⁶

$$\frac{\kappa}{q(\theta)} = \frac{p - w}{1 - \beta(1 - \delta)}, \quad (2)$$

where κ is the per-period vacancy posting cost, $q(\theta)$ is the vacancy filling probability (which falls in θ , serving as the equilibrating variable), p is labor productivity, w is wage, and β is the discount factor.⁷

Congestion in Hiring. The standard assumption in the DMP model is that productivity p is unaffected by the amount of hiring. By contrast, our paper explores a notion of diminishing returns in the amount of new hires. In essence, our quantitative model in Section 3 considers a richer version of the following aggregate production function:

$$Y = z(n_{\text{new}}^\sigma + n_{\text{inc}}^\sigma)^{\frac{1}{\sigma}}, \quad (3)$$

where z denotes total factor productivity (TFP), n_{new} denotes the mass of new hires, n_{inc} denotes the mass of incumbent workers, and σ governs the elasticity of substitution between new hires and incumbent workers (different hiring cohorts) in production.⁸ The standard model is nested with $\sigma = 1$, where new hires and incumbents are perfect substitutes.

The marginal product p of a newly hired worker is then given by:

$$p = z(n_{\text{new}}^\sigma + n_{\text{inc}}^\sigma)^{\frac{1}{\sigma}-1} n_{\text{new}}^{\sigma-1} = \frac{Y}{N} \cdot \frac{s_{\text{new}}^{\sigma-1}}{s_{\text{new}}^\sigma + s_{\text{inc}}^\sigma}, \quad (4)$$

where $N = n_{\text{new}} + n_{\text{inc}}$ is aggregate employment, and $s_{\text{new}} = n_{\text{new}}/N$ and $s_{\text{inc}} = n_{\text{inc}}/N$ are, respectively, the shares of new and incumbent workers in employment. Specifically, the marginal product of a new hire is equal to the average labor productivity times a congestion term that depends on the share of new hires in employment. Importantly, in the standard model with $\sigma = 1$, productivity is unaffected by the amount of hiring. By contrast, whenever $\sigma < 1$ (i.e., new hires and incumbents are imperfect substitutes in production), the marginal product of a new hire is decreasing in the share of new hires in employment—which we call *congestion in hiring*.

Definition (Congestion in Hiring). When $\sigma < 1$, the marginal product of a new hire is decreasing in the share of new hires in employment, $\partial p / \partial s_{\text{new}} < 0$.

Our proposed theory of countercyclical congestion amplifies new-hire productivity—and hence unemployment—fluctuations under two conditions: (i) the share of new hires in employment is

⁶In an important earlier paper, Coles and Moghaddasi Kelishomi (2018) modify the free-entry condition to be finitely elastic, and obtain improved quantitative performance too; we similarly discuss congestion via convex (UE) hiring costs as such an alternative source of congestion. Our production-based mechanism by contrast leaves free entry intact.

⁷For tractability, we ignore the potential dependence of wages on labor market tightness in this simple exposition.

⁸Our general production function in Section 3 also contains type-specific productivity shifters allowing us to remove any mechanical compositional effects across types.

countercyclical, and (ii) expansions in UE hiring push down the marginal product of new hires, so that increases in unemployment are not easily absorbed by new job creation. We now provide empirical evidence consistent with both ingredients.

2.2 Countercyclical Hiring Out of Unemployment, and the Role of Separations

We start by highlighting the empirical fact that UE flows are indeed countercyclical, and we link this cyclicity to the countercyclicity of job separations.

The Countercyclicity of New-Hire Share in the Data. Figure 1 present our main empirical fact and the first ingredient for countercyclical congestion: during recessions and in their aftermath, the ranks of the employed shift toward workers recently hired out of unemployment. Panel (a) presents raw levels, and Panel (b) presents log deviations from trend (using an HP-filter with a smoothing parameter of 100). We construct this measure using the 1976–2019 Current Population Survey (CPS) March Supplement (ASEC), which contains information on the number of weeks the respondent spent unemployed during the previous calendar year. We lead this annual time series by a year to align its reference period with the measurement of the unemployment rate, also ensuring consistency with the worker flow analysis we conduct subsequently. The time series show that this fact is robust to only counting unemployment spells longer than four weeks, and long-term unemployment totaling at least 26 weeks (after which recalls are essentially zero, see, e.g., Katz and Meyer, 1990; Fujita and Moscarini, 2017). The panels also include the log deviation of unemployment rate from its trend as an indicator of the business cycle.

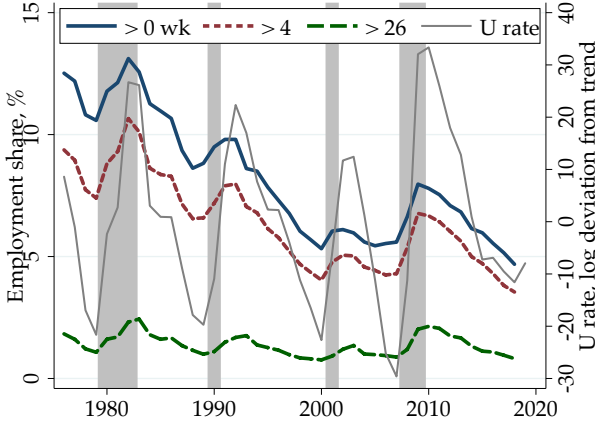
The Role of Countercyclical Separations. To understand the flow origins of countercyclical new-hire shares, we start from the definition of new hires as the product of job seekers and the job finding probability, $UE = f \cdot u$. Here, we draw on monthly CPS data covering 1976m1–2019m12. We track individuals switching their labor force status from one month to the next using the rotating-panel structure of the CPS. We construct quarterly averages of the monthly transition rates and only for visual clarity smooth the time series by taking four-quarter centered moving averages (but we use the underlying quarterly data for any statistic we report). Our approach follows Fujita and Ramey (2006) and we, therefore, relegate further details about data construction to Appendix A, including showing consistency of our time series to those reported in Shimer (2012), discussing correction for time-aggregation bias, considering nonemployment rather than unemployment alone, detrending with alternative smoothing parameters, and to considering other OECD countries.

Figure 2 Panel (a) shows that UE flows *increase* in recessions.⁹ This countercyclicity emerges if spikes in separations raise unemployment u strongly enough to offset the fall in job finding rates f . Formally, the (steady-state) elasticity of UE flows with respect to the unemployment rate is

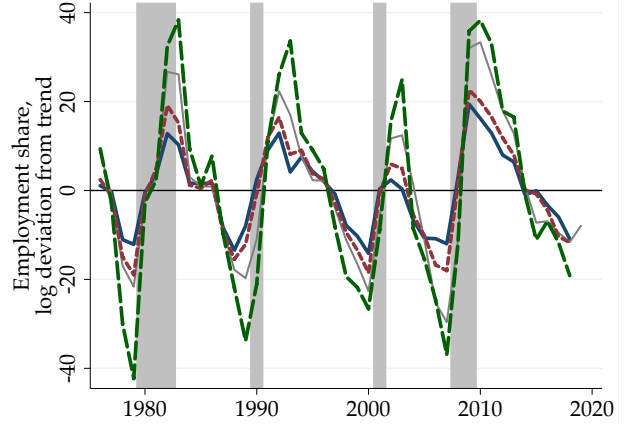
⁹The fact holds across OECD countries; we present the OECD evidence in Appendix A.5. This fact has been documented (but not studied as a source of amplification) by, e.g., Blanchard and Diamond (1990); Burda and Wyplosz (1994); Fujita and Ramey (2009); Elsby, Hobijn, and Şahin (2013).

Figure 1: Countercyclicity of New Hires

(a) Employment Shares of Workers with Unemployment Last Year by Total Weeks, and Unemployment Rate



(b) Cyclicity: Log Deviations from Trend



Notes: Panel (a) plots the share of employed workers who have undergone unemployment in the preceding calendar year for different amount of weeks (total). Panel (b) plots their log deviations from trend. The time series are HP filtered with a smoothing parameter of 100. Shaded regions denote NBER-dated recessions. Source: CPS March Supplement.

$\varepsilon_{UE,u} = \frac{df/f}{du/u} + 1 = \left((1-u) \left[-1 + \frac{d\delta/\delta}{df/f} \right] \right)^{-1} + 1$, and hence its sign is a priori ambiguous: a constant separation rate (as commonly assumed in models) imply *procyclical* UE flows; if separations are sufficiently countercyclical (i.e., if $\frac{d\delta/\delta}{df/f} < -\frac{u}{1-u}$), UE flows turn *countercyclical*.

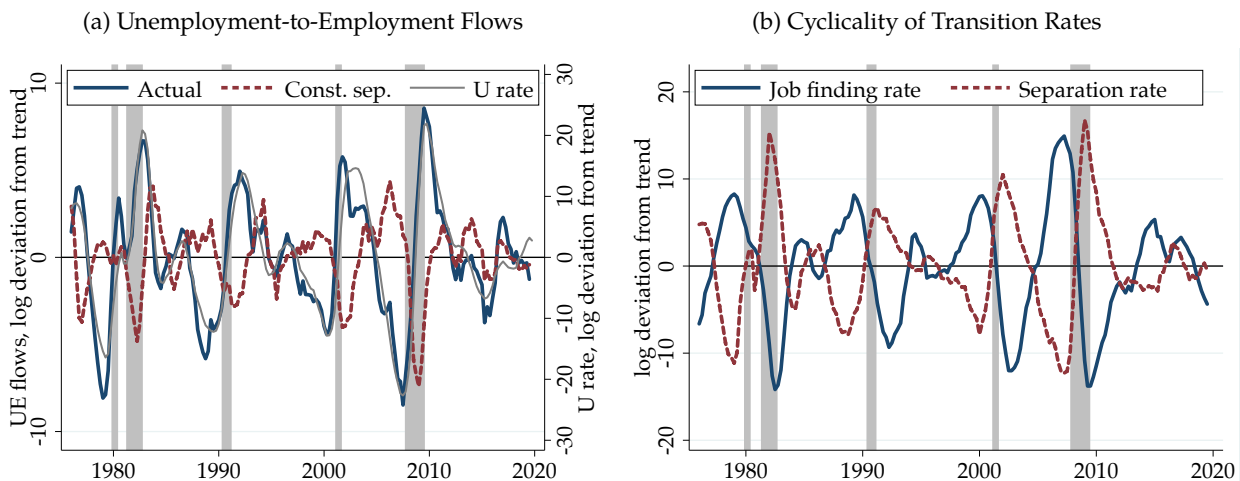
Figure 2 Panel (b) documents that in the US, the separation rate is indeed sufficiently countercyclical to generate countercyclical UE flows. In particular, we plot the detrended time series of both the job finding f and job separation δ rates, where we measure f and δ as the share of unemployed and employed workers in month $t-1$ who are observed to be employed and unemployed in month t , respectively.¹⁰ Their correlation is strongly negative at -0.717 and both time series are also relatively volatile, with standard deviations of 0.070 and 0.068 , respectively. Therefore, in the US data, the condition for countercyclical UE flows is safely satisfied $\frac{d\delta/\delta}{df/f} \approx -0.698 < -\frac{u}{1-u} \approx -0.067$.

Finally, Panel (a) further highlights the importance of separations in generating countercyclical UE flows by plotting a counterfactual UE time series implied by a constant job separation rate.¹¹ In the absence of separation rate movements, UE flows indeed become *procyclical* (their correlation with unemployment is -0.389 rather than 0.802 as in the data). Appendix Figure A1 traces out these elasticities as scatter plots of time series, for UE flows and the new-hire share.

¹⁰In the data, and later on in the model, we specify discrete-time transition probabilities while using the conventional term “rates” interchangeably. Appendix A.1 presents details on the measurement of job finding and separation rates. Appendix A.2 replicates our empirical analysis for alternative treatments of these transition rates.

¹¹Specifically, we iterate on the law of motion for unemployment given by $\tilde{U}_{t+1} = (1-f_t)\tilde{U}_t + \delta(L_t - \tilde{U}_t)$ to construct the counterfactual time series for the *mass* (rather than rate) of unemployed \tilde{U}_t over our sample, where f_t and L_t denote the observed job finding rate and labor force in month t , and $\bar{\delta}$ denotes the sample average of the job separation rate. Then our counterfactual time series for UE flows is $\tilde{U}E_t = f_{t-1}\tilde{U}_{t-1}$.

Figure 2: Unemployment-to-Employment (UE) Flows and Worker Transition Rates



Notes: Panel (a) plots the log deviations in quarterly averaged monthly UE flows and the counterfactual flows implied by a constant EU rate set to its sample mean. Panel (b) plots log deviations of quarterly-averaged monthly UE and EU rates from their trends. All time series are HP filtered with a smoothing parameter of 1,600 and smoothed by taking centered four-quarter moving averages for visual clarity. Shaded regions denote NBER-dated recessions. Source: CPS monthly files.

UE vs. Total Hires (Including Job-to-Job Transitions). While UE flows are countercyclical, job-to-job transitions (and quits) drop dramatically in recessions (see, e.g., Mercan and Schoefer, 2020), so that total (rather than those only out of unemployment) hires are not countercyclical. Our model studies countercyclical congestion in jobs filled by workers hired out of unemployment, their share in employment, and (their effect on) flows between unemployment and employment, and hence thinks of hires from unemployment and job-to-job movers as filling different types of jobs.

2.3 Congestion in Hiring: Evidence from Separation Shocks

We now provide evidence consistent with the second ingredient necessary for congestion to generate amplification: in the data, hiring does not absorb sudden increases in unemployment as much as predicted by the standard DMP model, which in our model is rationalized by a diminishing product of new hires. To this end, we derive a key prediction of congestion that we then empirically support: innovations to separations cause a drop in labor market tightness.

Model Prediction. To formulate the model prediction and to dissect the amplification arising from congestion, we derive the on-impact elasticity of labor market tightness to a one-time increase in (last period's) separation rate, δ_{-1} . We substitute the expression for new-hire productivity in Equation (4) into the job creation condition in Equation (2) and make use of the fact that $n_{new} = u f(\theta)$ to get:

$$\frac{\kappa}{q(\theta)} = \frac{\beta}{1 - \beta(1 - \delta)} \left(D (u f(\theta))^{\sigma-1} - w \right), \quad (5)$$

where D is a positive constant.¹² We consider the impact of an infinitesimal increase in (last period's) separation rate (for a fixed wage w) which increases the unemployment pool, but leaves other fundamentals (in particular TFP) unchanged. For concreteness, and as is common in DMP models, we assume that the job filling and job finding probabilities are given by $q(\theta) = \bar{m}\theta^{-\mu}$ and $f(\theta) = \bar{m}\theta^{1-\mu}$, respectively, where $\bar{m} > 0$ captures the efficiency of the matching process and $\mu \in (0, 1)$ is the elasticity of the matching function with respect to vacancies. Using the law of motion for unemployment in Equation (1) and implicitly differentiating θ with respect to δ_{-1} in Equation (5) we get

$$\varepsilon_{\theta, \delta_{-1}} = \frac{(\sigma - 1)p}{(p - w)\mu - (\sigma - 1)p(1 - \mu)} \varepsilon_{u, \delta_{-1}}, \quad (6)$$

where $\varepsilon_{u, \delta_{-1}} = \delta_{-1}(1 - u_{-1})/u > 0$. Equation (6) clarifies the distinction between the standard DMP model ($\sigma = 1$) and our framework with congestion in hiring ($\sigma < 1$). When $\sigma = 1$, we have $\varepsilon_{\theta, \delta_{-1}} = 0$, echoing the irrelevance of market size in the standard DMP model for labor market tightness. This happens because—in the absence of a change in productivity—vacancies simply scale one-for-one with unemployment to satisfy the free-entry condition in Equation (2). By contrast, in our congestion model with $\sigma < 1$, $\varepsilon_{\theta, \delta_{-1}}$ is negative: labor market tightness falls as the larger unemployment pool brought upon by the separation shock implies a larger cohort of new hires out of unemployment. This swelling of the ranks of new hires, in turn, reduces their marginal product and—compared to a model without congestion—depresses incentives for further job creation.

Empirical Test. We now test the prediction in Equation (6): innovations to separations lead to drops in labor market tightness, in contrast to the zero effect in the standard model without congestion. Specifically, using a VAR model, we study the response of labor market tightness to a separation shock. We focus on the following vector of variables:¹³

$$\mathbf{y}_t = [\ln ALP_t, \ln \delta_t, \ln \theta_t],$$

where ALP is average labor productivity (measured as output per worker in the non-farm business sector), δ is the separation rate (EU flows divided by beginning-of-period employment), and θ denotes labor market tightness (vacancies from Barnichon, 2010, divided by unemployment). To be consistent with our subsequent quantitative analysis and due to data limitations (ALP is measured on a quarterly frequency), we convert the monthly job separation rate to a quarterly measure.

¹²For simplicity we assume that the first terms in Equation (4) are constant and define $D = z(n_{\text{new}}^\sigma + n_{\text{inc}}^\sigma)^{\frac{1}{\sigma}-1}$.

¹³Coles and Moghaddasi Kelishomi (2018) also study the response of the labor market to separation shocks. Their modification of the free-entry condition also constitutes a congestion mechanism, and resembles our alternative specification of congestion via convex (UE) hiring costs, which, however, misses some of our key results pertaining to wage fluctuations and earnings losses from displacement.

We then estimate the following VAR model:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}(L)\mathbf{y}_{t-1} + \mathbf{v}_t, \quad (7)$$

where \mathbf{c} is a constant term, $\mathbf{A}(L)$ is a lag polynomial, and $\mathbf{v}_t \sim (0, \mathbf{\Omega})$ is a vector of error terms with mean zero and variance-covariance matrix $\mathbf{\Omega}$. We include four lags of the endogenous variables in our specification and identify productivity and separation shocks using a recursive identification scheme (or, equivalently, using a Cholesky decomposition of $\mathbf{\Omega}$). Our timing assumption is that *ALP* has a contemporaneous effect on both δ and θ . In contrast, δ only has a contemporaneous effect on θ and, finally, θ affects the endogenous variables only with a lag. We then study impulse responses to an orthogonalized shock to δ , to isolate the effect of movements in job separations from that of productivity fluctuations.¹⁴ In Appendix B.2 we provide an extensive discussion of potential issues with our VAR approach to identifying congestion and we conclude that the leading candidates of identified shocks in the literature are unlikely to confound our estimation of congestion dynamics. Finally, we use the innovations in the separation rate to discipline our model, but recognize that these innovations are outcomes of endogenous separations in response to unmodeled structural shocks that shift job surplus.

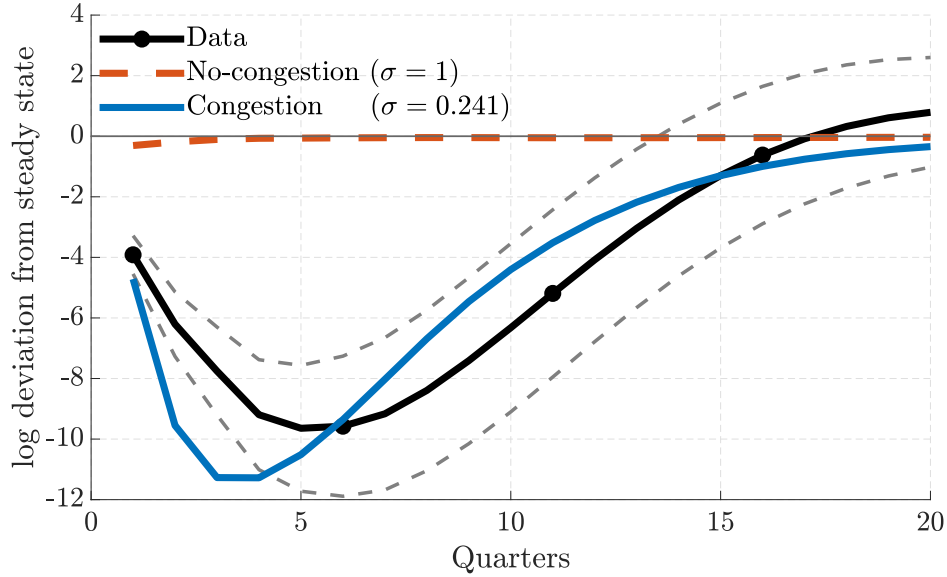
Figure 3 plots the empirical IRF of labor market tightness to an innovation in the separation rate. The empirical response of labor market tightness is significantly negative and persistent, and hence clearly rejects the insensitivity of labor market tightness, which the figure also plots as a benchmark (formally presented in Section 4). By contrast, and consistent with our notion of congestion ($\sigma < 1$), firms expand vacancies much less relative to the standard framework, and hence fail to absorb the newly unemployed workers quickly. In our quantitative model, we pin down the degree of congestion by having our model match this IRF, which the figure foreshadows. Relatedly, the companion Appendix Figure A11 reports the empirical IRF of unemployment rather than labor market tightness, indicating a much slower recovery in unemployment, due to the drop in the job finding rate.

2.4 Existing Evidence and Potential Sources of Congestion in Hiring

Besides the *aggregate* time-series evidence presented above, we now discuss the considerable amount of supportive *cross-sectional* evidence pointing to congestion in hiring as well as its concrete potential channels.

¹⁴This orthogonality with productivity holds exactly in the first period. In Appendix Figure A12, we present the IRFs of ALP to the δ shock. Importantly, if anything, the empirical process indicates (insignificantly) positive ALP responses to a positive separation rate shock in the transition periods. Hence, the comovement of productivity with the separation shock would lead to an increase rather than decrease of labor market tightness (and a decrease in unemployment). Moreover, evidence suggests that the composition of the unemployment pool improves and that firms find it profitable to increase their hiring standards in recessions (see, e.g., Mueller, 2017; Modestino, Shoag, and Ballance, 2016). Congestion arises in our model as long as the pool of the unemployed differ from the employed.

Figure 3: Impulse Response of Labor Market Tightness to a Separation Shock



Notes: The figure plots the empirical response of labor market tightness to a separation shock (dashed lines are one standard deviation confidence bands), together with model implied responses. “No-congestion ($\sigma = 1$)” model refers to the standard model with homogeneous workers. “Congestion ($\sigma = 0.241$)” model refers to our model under the preferred calibration. The model IRFs are explained in Section 4.

Cross-Sectional Evidence on Congestion in Hiring. First, at the firm level, Doran, Gelber, and Isen (2020) show that one hire randomly assigned across firms (leveraging US visa lotteries) crowds out (more than) one additional subsequent hire into that job type—which would imply full congestion at the firm level, and a target employment count in narrowly defined entry level job types. Such hiring targets are also consistent with qualitative evidence implying that entry-level jobs are imperfect substitutes for higher-tier jobs (see, e.g., the “ports of entry” and internal labor markets described in Doeringer and Piore, 1985).

Second, as shown in the meta analysis in Mercan and Schoefer (2020) covering 15 studies, local labor markets appear to absorb spikes in unemployment sluggishly, as short-run employment spillovers from firms directly affected by labor demand shifters onto peer firms appears limited. For example, employment subsidies targeting some eligible firms have no or strikingly limited effects on hiring by ineligible employers in the same local labor market (Cahuc, Carcillo, and Le Barbanchon, 2018; Giupponi and Landais, 2020). Similarly, sharp labor demand reductions and mass layoffs by particular plants or sectors, which closely approximate a separation shock that leaves peer firms’ job values constant, do not lead other employers to expand even in the same industry or in other tradable industries in the short run (e.g., Mian and Sufi, 2014; Gathmann, Helm, and Schönberg, 2018).

Potential Concrete Channels of Congestion in Hiring. We now discuss potential concrete mechanisms that may generate congestion in hiring, where our model—taken literally—assumes that

cohorts of new hires enter the aggregate production function as imperfect substitutes to incumbent workers (older cohorts of hires).

First, a widely studied source of difference between new hires and incumbent workers is general or firm-specific human capital accumulation on the job. Workers may partially lose such skills during unemployment, consistent with evidence for large earnings losses upon job displacement (Jacobson, LaLonde, and Sullivan, 1993), unemployment spells lowering applicants' attractiveness for employers (Kroft, Lange, and Notowidigdo, 2013), and with turbulence models of unemployment (Ljungqvist and Sargent, 1998, 2004; den Haan et al., 2005; Kehoe et al., 2019).

Second, a long literature in personnel economics offers another reason for why new hires and incumbent workers may be imperfect substitutes: internal labor markets (ILMs). ILMs feature a notion of careers, i.e., a progression from limited entry-level jobs ("ports of entry") to higher-tier jobs, which are predominantly filled by incumbents (see, e.g., Doeringer and Piore, 1985; Lazear and Oyer, 2004). ILMs have rigid structures, with the employment distribution across job levels and career progressions remaining stable within a firm even in the face of substantial employment growth (see, e.g., Baker, Gibbs, and Holmstrom, 1994)—thereby curbing a quick expansion of hiring into entry-level jobs.¹⁵

Third, rather than firm-level processes, imperfect substitution between new and incumbent workers may work through the quality of hiring firms. Workers hired in recessions appear to switch—specifically downgrade—occupations (see, e.g., Altonji, Kahn, and Speer, 2016; Huckfeldt, 2016), with firms upgrading skill requirements for new hires (see, e.g., Modestino, Shoag, and Ballance, 2016; Carrillo-Tudella, Gartner, and Kaas, 2021). In recessions, low-productivity and low-wage employers expand hiring relatively more strongly and, therefore, absorb larger shares of new hires (see e.g. Schmieder, von Wachter, and Heining, 2019; Oreopoulos, von Wachter, and Heisz, 2012). A complementary literature studies the reallocation of hires by various margins across heterogeneous firms (see, e.g., Moscarini and Postel-Vinay, 2012; Haltiwanger, Hyatt, Kahn, and McEntarfer, 2018). Here, we caveat that our aggregate production function (and the congestion parameter σ it features) is likely difficult to relate to micro firm behavior and unlikely to accurately capture lifecycle or episodic growth events of individual firms.

Fourth, convex costs in (UE) hiring may underlie congestion. For instance, firms may need to create new jobs and reorganize production (see, e.g., Fujita and Ramey, 2007; Coles and Moghadasi Kelishomi, 2018; Mercan and Schoefer, 2020), screen applicants (see, e.g., Hall, 2005a; Engbom, 2020), and then train the new hires (see, e.g., Silva and Toledo, 2009). To the degree that these adjustments disrupt production by, e.g., moving incumbent workers to training purposes (as in, e.g., Faccini and Yashiv, 2020), the deep sources of a convex hiring cost may still reflect production function features. While we show robustness of our results to capturing congestion using

¹⁵The canonical case study of a mid-sized US firm in Baker, Gibbs, and Holmstrom (1994) documents some evidence for ports of entry, with 99% such jobs being filled with outside hires, who then move up the internal career ladder, with 75% of higher-level jobs filled through internal career moves. Moreover, they document that "almost everyone who enters the firm at the lower levels goes through the same number of promotions before reaching higher levels" (p. 915) and that this structure is rigid since "the firm added no new levels over time as employment tripled. Major titles in 1969 were major titles in 1988, with little change in the distribution of employment across titles" (p. 916).

convex hiring costs, there exists an important difference between the two modelling choices. In particular, congestion through costs cannot explain the observed cohort-specific wage profile differences (without other model adjustments such as staggered wage bargaining as opposed to period-by-period wage bargaining) that we discuss in the next sections.

3 A Search Model with Countercyclical Congestion

We now integrate countercyclical congestion into an otherwise standard DMP model, providing a full and quantitative treatment of the mechanism sketched in Section 2. In Section 4 we calibrate the model and in Section 5 we study its quantitative performance for core labor market variables. Section 6 shows that our framework provides a unified explanation for a range of other labor market patterns connected to unemployment fluctuations that have been difficult to rationalize within a single framework.

To model countercyclical congestion in hiring from unemployment, we add two ingredients into the canonical DMP framework. First, we generate countercyclical UE flows by adding countercyclical job separations. Second, to obtain congestion, our model features an aggregate production function with diminishing returns in new hires—arising from imperfect substitution between hiring cohorts. When UE flows rise, as they do in recessions, new hires become relatively abundant among the employed. The marginal product of new hires falls, rationalizing why firms do not absorb laid off workers as quickly as predicted by no-congestion models.

3.1 Worker Heterogeneity: Cohort-Specific Types and Congestion

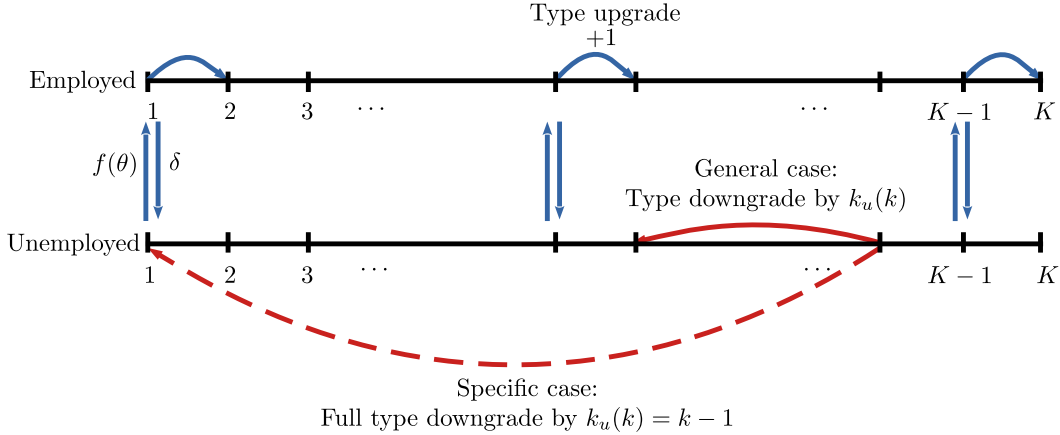
We begin by describing the key extension of our model: worker heterogeneity that depends on time since hiring, and their imperfect substitutability in production. This feature generates diminishing returns in new hires, which acts as the source of congestion in our model.

Worker Types. Workers are heterogeneous in their type $k \in \mathcal{K} = \{1, \dots, K\}$, with maximum type $K \geq 1$. Index k stands for various economic mechanisms whereby workers with different labor market histories become different from the point of view of employers. Section 2 previewed a simple model where $K = 2$.

Figure 4 summarizes how worker types evolve in our setting during employment and unemployment spells. Each period a worker is employed, she moves up by one level, i.e., $k_{t+1} = k_t + 1$, where t indexes time. While unemployed, workers downgrade by $k_u(k)$ steps, i.e. $k_{t+1} = k_t - k_u(k_t)$, where $k_u(k) \in \{0, 1, \dots, k-1\}$ determines the size of the downgrade as a function of current type k . This setup nests various possibilities ranging from no downgrading $k_{t+1} = k_t$, achieved by setting $k_u(k) = 0$, to full downgrading to $k_{t+1} = 1$ for all types k , achieved by setting $k_u(k) = k - 1$.

This tractable and agnostic type evolution has two key advantages over modeling multiple concrete mechanisms—discussed in Section 2—that could underlie it. First, it allows for a direct

Figure 4: Worker Type Evolution



comparison to the standard DMP model, which our framework nests. Second, it allows us to estimate the *overall* impact of countercyclical congestion on unemployment fluctuations, independent of its sources.

Congestion: Production with Diminishing Returns to Worker Types. Worker heterogeneity matters through the aggregate production function. Workers of different types produce intermediate goods using a linear technology converting one unit of labor to a unit of intermediate good differentiated by worker type. We denote the stock of type- k workers (and hence intermediate inputs) by $\{n_k\}_{k=1}^K$. Intermediate inputs are sold to a final good producer in a competitive market at prices $\{p_k\}_{k=1}^K$. The final good producer combines these inputs into a final consumption good (the numeraire). Final good production is subject to fluctuations in TFP z . The aggregate production function is a generalization of the one in (3) and is given by

$$Y = z \left(\sum_{k=1}^K \alpha_k n_k^\sigma \right)^{1/\sigma}, \quad (8)$$

where α_k is a type-specific productivity shifter associated with worker type k , and σ governs the elasticity of substitution between inputs. This functional form exhibits overall constant returns to scale and a constant elasticity of substitution across worker types equal to $\frac{1}{1-\sigma}$.¹⁶ The standard DMP model is nested as a special case when worker types are perfect substitutes for one another (and no differences in productivity weights α_k), permitting us to isolate the congestion mechanism.

The competitive price for each intermediate input k reflects the marginal product of labor-type

¹⁶In Appendix C, we present a generalization that allows for perfect substitution between subsets of worker types, thereby permitting one to generalize the type upgrade and downgrade processes further.

k engaged in that good's production:

$$p_k = \alpha_k n_k^{\sigma-1} \frac{Y}{\sum_{l=1}^K \alpha_l n_l^\sigma} = \alpha_k s_k^{\sigma-1} \frac{1}{\sum_{l=1}^K \alpha_l s_l^\sigma} \frac{Y}{N}, \quad (9)$$

where $N = \sum_{l=1}^K n_l$ denotes aggregate employment, Y/N is average labor productivity (ALP), and $s_l = n_l/N$ denotes the employment share of type- l workers. Equation (9) makes clear that the productivity of a given worker type features diminishing returns in its *employment share*.

Specific Case: Full Downgrading to $k = 1$ Upon Job Loss. In the specific case that upon job loss, workers fully downgrade to $k = 1$ (i.e., $k_u(k) = k - 1$ for all k), all unemployed workers become the same type. Hence, all UE hires are the same type, and climb the worker-type ladder as one cohort. This case permits an easy representation of new hires' marginal product of labor, namely $p_{k=1}$.

Figure 5 Panel (a) traces out the relationship between the marginal product of new hires p_1 against their employment share under the assumption of full type downgrading. We plot this relationship for two levels of the congestion parameter $\sigma \in \{0.241, 1\}$. To isolate the influence of worker heterogeneity on congestion from mechanical composition effects (e.g., Mueller, 2017; Ferraro, 2018; Hagedorn, Manovskii, and Stetsenko, 2016), we set α_k such that steady-state marginal products are normalized to one for all worker types, i.e., $p_k = 1$ for all k , and for each σ level.

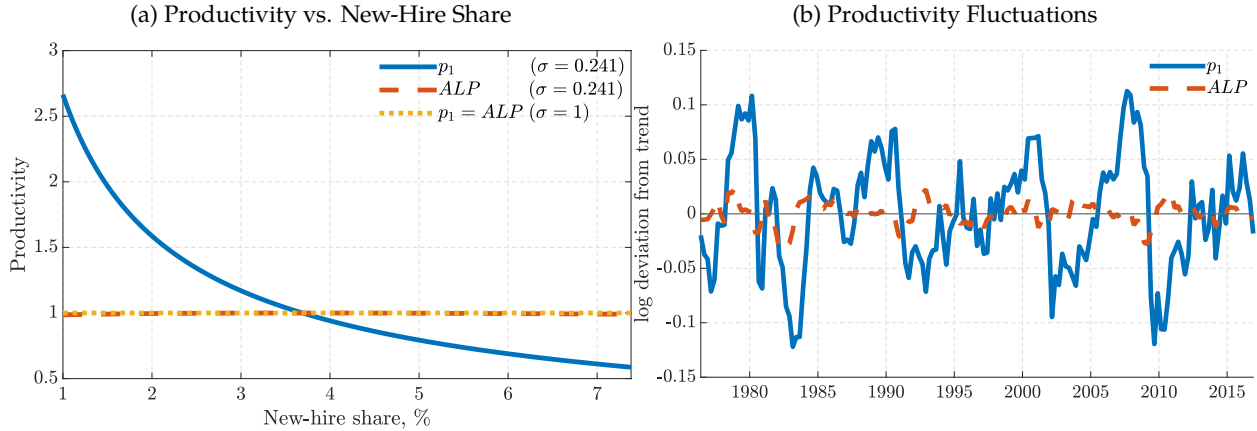
The flat yellow dotted line captures the case of $\sigma = 1$, for which workers are perfect substitutes, and each type's marginal product simply equals the average labor productivity, Y/N . Shifts in the share of new hires have no effect on productivity. This specification renders the model isomorphic to the standard model with homogeneous workers and no congestion in hiring.

If $\sigma < 1$, the economy exhibits diminishing returns in each type k . We set $\sigma = 0.241$, foreshadowing our estimate for congestion in Section 4.3. The blue solid line is the productivity of new hires, which falls sharply when new hires become abundant. Specifically, a 10% increase in the share of new hires (that is, 0.4ppt off the baseline of 4%) lowers productivity by around 7.6% (the local slope of $1 - 0.241 = 0.759$). As another way to judge σ , the implied elasticity of substitution of worker types (cohorts) is around $1/(1 - \sigma) \approx 1.3$.

Importantly, these movements in new-hire productivity have no visible effect on the naive ALP concept Y/N (red dashed line), which is essentially flat, even for large changes in hiring. This property is due to the CRS-CES production function. Therefore, the large fluctuations in productivity of new hires that our model implies can be masked by—and hence be consistent with—the smooth ALP in the data.

Implications for Productivity Fluctuations. Our congestion mechanism implies large fluctuations in new-hire productivity compared to ALP. Figure 5 Panel (b) plots the time series (log deviations from trend) of productivity of new hires p_1 , along with the average labor productivity Y/N . We construct new-hire productivity p_1 by feeding in the observed share of UE hires, $s_{1,t}$, at each quarter, which gives $p_{1,t} = \alpha_1 s_{1,t}^{\sigma-1} ALP_t \frac{1}{\sum_{k=1}^K \alpha_k s_{k,t}^\sigma}$, where ALP_t is the observed average labor

Figure 5: Flow Productivity and The Size of the Hiring Cohort



Notes: Panel (a) plots the marginal product of new hires and average labor productivity as a function of the employment share of new hires for different values of congestion parameter σ . Steady-state average labor productivity and each type's marginal product are normalized to one for both calibrations of σ . Panel (b) plots the empirical US time series for average and new-hire productivities. Both time series are in logs and detrended using an HP-filter with a smoothing parameter of 1,600.

productivity.¹⁷ At $SD(p_1) = 0.052$, the volatility of new-hire productivity is essentially *five* times as high as that of the standard average labor productivity ($SD(ALP) = 0.010$) typically used in the literature as a driving force (e.g., Shimer, 2005; Hall, 2005b; Hagedorn and Manovskii, 2008; Pissarides, 2009). Yet, over the US business cycles, the maximum amplitude of new-hire productivity remains tightly within an interval of plus and minus 10% around its trend.

Discussion: Alternative Type Processes. When specifying the full model below, we present the general case regarding type downgrading, and then calibrate our model under the specific assumption of full type downgrading. We also show robustness to alternative type-downgrading specifications; once recalibrated to match the same congestion targets, these variants turn out to be *isomorphic*.

Discussion: Segmentation of Cohorts. The assumption that hiring cohorts remain segmented throughout their tenures, even, e.g., 20 years into the job, may appear unappealing. However, high turnover rates in the US economy wash out cohort effects. For instance, fewer than 5% of workers remain in the same job for 20 years. Moreover, if in reality congestion occurred only early in the job (i.e., a lowering of K), the calibrated model would simply require a larger degree of diminishing returns, i.e., a lower σ , to match the empirical congestion response in hiring. In addition, the new-hire productivity time series exhibits persistence, which compresses productivity differentials

¹⁷For this exercise (but not in subsequent analyses), we ignore fluctuations in the third term arising from the history of the law of motion of worker types, which are small but would otherwise force us to drop the first 160 quarters in our sample if we followed our eventual specification of $K = 160$. We therefore consider at each point the deviations from steady state in only the new-hire share while ensuring that the shares of other types for $k > 1$ drop accordingly.

between *adjacent* cohorts. Finally, Appendix G shows that our results are robust to the maximum number of cohorts K .

3.2 Environment and Timing

Except for worker heterogeneity and the associated aggregate production function described above, the remainder of the model follows the standard DMP model as in, e.g., Shimer (2005).

Environment. There is a continuum of workers comprising the labor force of mass L . They are infinitely lived and ex-ante identical. Preferences are risk-neutral, with discount factor $\beta \in (0, 1)$. Individuals own the two types of producers: intermediate-input producers (“firms”), which use labor to produce output that they sell in a perfectly competitive market to a final good producer. The latter “retailer” bundles the intermediate goods into a final consumption good using the technology in Equation (8). The retailer pins down intermediate input prices, which stand for the marginal products of worker types.

Matching. The labor market is subject to search frictions. Jobs take the form of single worker-firm matches and produce intermediate goods using a linear technology. Meetings between unemployed workers and vacancies (firms with unfilled jobs) are random, and follow a constant-returns-to-scale matching function $M(u, v) < \min\{u, v\}$, where u is the mass of unemployed searching for jobs and v is the mass of open vacancies. Labor market tightness is the ratio of vacancies v to unemployment u , $\theta = v/u$. The job finding rate for an unemployed worker is $f(\theta) = \frac{M}{u} = M(1, \theta)$; the vacancy filling rate for a firm is $q(\theta) = \frac{M}{v} = M(1/\theta, 1)$.

Separations. Each period, active matches separate with exogenous but time-varying rate δ . These separations are an ad-hoc event rather than arising from endogenous decisions between the worker and firm in response to shocks to surplus. We take this route, as we conjecture that modeling endogenous separations should leave our key results intact provided such an extended model matches the impulse response of labor market tightness to separation shocks unrelated to productivity movements (as documented in Figure 3).¹⁸

No Job-to-Job Transitions. We primarily refer to the k -types as denoting skills gained on the job. Some of this upgrading may also reflect the progression of a worker through the original employer’s internal labor market. In the broadest sense, one could think of the process as incorporating even job ladders involving employer switches, but we do not explicitly model such employer-to-employer transitions for simplicity and because our ultimate interest is in hiring out of unemployment. Informally, we think of job-to-job transitions as leaving workers on track in terms of their type evolution. The crucial feature our model requires is that the type evolution

¹⁸An interesting question beyond the scope of our model with exogenous separations is whether endogenous separations become harder to justify if skill loss is involved (see den Haan, Haefke, and Ramey, 2005, for a discussion).

when employed is different from that in unemployment as in models of turbulence (Ljungqvist and Sargent, 1998, 2004). Hence, our focus and notion of a job echoes the concept of “employment cycles” uninterrupted by unemployment spells and potentially including job-to-job transitions as in Hagedorn and Manovskii (2013).¹⁹

Aggregate State Variables. The economy is subject to aggregate shocks, namely to the job separation rate δ and to TFP in final good production z . Additional state variables are the worker distributions across k types in unemployment (due to random search) and over employment (due to the CES production function). Below, we index value functions and variables by time subscript t , which, besides time, implicitly captures all the relevant aggregate state variables.

Timing. At the beginning of each period, aggregate productivity z and separation rate δ are realized. Worker-firm matches (both those active last period and those formed last period) are destroyed at rate δ , in which case the workers become unemployed. The surviving matches produce the intermediate inputs differentiated by the type of the worker k , which the retailer bundles into the final consumption good. Workers consume their wage or unemployment benefits, depending on their employment status and k -type. Employed workers upgrade by one type, and unemployed workers downgrade by $k_u(k)$ types. The period closes by the search stage. Firms post vacancies and unemployed workers search for jobs, which determine market tightness. New matches are formed.

Evolution of Type Distributions. The worker distributions over types evolve according to the following laws of motion:

$$\begin{aligned} u_{k-k_u(k),t} &= (1 - f(\theta_{t-1})) u_{k,t-1} + \delta_t e_{k-k_u(k),t} \text{ for all } k \\ e_{k-k_u(k),t} &= (1 - \delta_{t-1}) e_{k-k_u(k)-1,t-1} + f(\theta_{t-1}) u_{k,t-1} \text{ for all } k, \end{aligned} \tag{10}$$

with $e_{k,t}$ denoting the beginning of period employment mass of type- k workers. The labor input that enters production is equal to $n_{k,t} = (1 - \delta_t) e_{k,t}$, as separations occur at the beginning of a period. Type-specific unemployment $u_{k,t}$ is written after the separation stage (but before type changes, which occur at the end of the period). Aggregate unemployment is given by $u_t = \sum_{k=1}^K u_{k,t} = L - \sum_{k=1}^K n_{k,t}$.

3.3 Worker and Firm Problems, and Equilibrium

We now describe the worker and firm problems, wage determination, the match surplus, and the labor market clearing condition, and we define the equilibrium of our model.

¹⁹If the mechanism worked through the job ladder only, then workers would have an incentive to search harder for the more-productive jobs in recessions. However, even in such a setting, the model would need to be consistent with the observed drop in labor market tightness following separation shocks.

Worker and Firm Problems. We cast the worker and firm problems recursively. The value functions are written as of the consumption/production stage within the period.

The value of an employed worker of type k is

$$W_{k,t} = w_{k,t} + \beta \mathbb{E}_t \left[(1 - \delta_{t+1}) W_{k+1,t+1} + \delta_{t+1} U_{k+1,t+1} \right], \quad (11)$$

where $w_{k,t}$ is the bargained real wage (to be described below), which the worker consumes. Next period, the worker keeps her job at rate $1 - \delta_{t+1}$ (realized at the beginning of the period) and otherwise becomes unemployed.

The value of an unemployed worker of type k is

$$U_{k,t} = b + \beta \mathbb{E}_t \left[f(\theta_t)(1 - \delta_{t+1}) W_{k-k_u(k),t+1} + (1 - f(\theta_t)(1 - \delta_{t+1})) U_{k-k_u(k),t+1} \right], \quad (12)$$

where b is the flow value of unemployment.²⁰ If the worker contacts a firm and does not separate at the beginning of the next period, she becomes employed. Otherwise the worker stays unemployed. Upon spending the current period in unemployment, the worker's type downgrades to $k - k_u(k)$, whether she finds a job or not.

Firm problems mirror that of the workers. The value of a vacancy is

$$V_t = -\kappa + \beta \mathbb{E}_t \left[q(\theta_t)(1 - \delta_{t+1}) \sum_k \frac{u_{k,t}}{u_t} J_{k-k_u(k),t+1} + (1 - q(\theta_t)(1 - \delta_{t+1})) V_{t+1} \right], \quad (13)$$

where the firm pays flow cost κ to maintain the vacancy and $\sum_k \frac{u_{k,t}}{u_t} J_{k-k_u(k),t+1}$ is the average job value from randomly meeting unemployed workers of different types k at time t .

A firm that employs a worker of type k has value

$$J_{k,t} = p_{k,t} - w_{k,t} + \beta \mathbb{E}_t \left[(1 - \delta_{t+1}) J_{k+1,t+1} + \delta_{t+1} V_{t+1} \right], \quad (14)$$

where $p_{k,t}$ is the price of the type-specific good produced by the match, taken as given by the firm. The firm pays the worker a bargained wage $w_{k,t}$. The match continues until the exogenous separation shock dissolves it.

Surplus, Wage Determination, and Free Entry. Total surplus from a match is the sum of worker and firm surpluses, and is given by

$$S_{k,t} = W_{k,t} - U_{k,t} + J_{k,t} - V_t. \quad (15)$$

²⁰We will interpret b , interchangeably, as unemployment insurance since extending the model with a government levying lump-sum taxes to finance such a policy leaves the rest of the model unchanged.

The individual value functions in Equations (11)–(14) and the definition of surplus in Equation (15) yield the following surplus value:

$$S_{k,t} = p_{k,t} - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1}) S_{k+1,t+1} - f(\theta_t)(1 - \delta_{t+1}) \phi S_{k-k_u(k),t+1} + U_{k+1,t+1} - U_{k-k_u(k),t+1} \right], \quad (16)$$

and the value of unemployment can be expressed in terms of match surplus as follows:

$$U_{k,t} = b + \beta \mathbb{E}_t \left[f(\theta_t)(1 - \delta_{t+1}) \phi S_{k-k_u(k),t+1} + U_{k-k_u(k),t+1} \right]. \quad (17)$$

The wage for worker type k is determined period-by-period by generalized Nash bargaining:

$$w_{k,t} = \arg \max (W_{k,t} - U_{k,t})^\phi (J_{k,t} - V_t)^{1-\phi}, \quad (18)$$

where $\phi \in (0, 1)$ is the bargaining power of the worker. Due to transferable utility, this bargaining problem implies linear surplus sharing rules given by

$$W_{k,t} - U_{k,t} = \phi S_{k,t} \quad \text{and} \quad J_{k,t} - V_t = (1 - \phi) S_{k,t}, \quad (19)$$

where the worker captures share ϕ of the total match surplus, and the firm captures the rest.

Free entry of firms pins down $V_t = 0$ for all t . Equation (13) therefore implies

$$\frac{\kappa}{q(\theta_t)} = \beta(1 - \phi) \mathbb{E}_t \left[(1 - \delta_{t+1}) \sum_k \frac{u_{k,t}}{u_t} S_{k-k_u(k),t+1} \right]. \quad (20)$$

Stochastic Equilibrium of the Congestion Model. The stochastic equilibrium of the model is a set of value functions for match surplus $\{S_k\}_{k=1}^K$ and unemployment $\{U_k\}_{k=1}^K$, intermediate input prices $\{p_k\}_{k=1}^K$, beginning-of-period masses of unemployed $\{u_k\}_{k=1}^K$ and employed $\{e_k\}_{k=1}^K$, end-of-period quantities of intermediate goods $\{n_k\}_{k=1}^K$, and labor market tightness θ , such that:

- match surplus S_k solves the Bellman equation in Equation (16) for all k ,
- unemployment value U_k solves the Bellman equation in Equation (17) for all k ,
- intermediate goods prices p_k satisfy Equation (9) for all k ,
- worker masses, u_k and e_k , follow the laws of motion in Equation (10) for all k ,
- end-of-period intermediate goods are given by $n_k = (1 - \delta)e_k$ for all k ,
- market tightness θ solves the free-entry condition in Equation (20),
- exogenous state variables z and δ follow stochastic processes specified in Section 4.2.

4 Model Parameterization

We now discuss our calibration strategy before turning to studying the model quantitatively in the following sections. We calibrate the model to match moments of the US economy, in the period covering 1976Q2–2019Q4 (except for vacancies and labor market tightness, for which the time series end in 2016, Barnichon, 2010). The model period is one quarter. We, therefore, convert our monthly transition rates to quarterly values and use the HP filter with a smoothing parameter of 1,600 to extract the cyclical component of simulated time series.²¹

Table 1 summarizes the model parameters and the targets we use to discipline them. Appendix D provides technical details for how we solve and simulate the model. Absent congestion, the model mirrors the standard DMP model, which we calibrate as in Shimer (2005). Therefore, we begin by describing a set of standard parameters commonly found in other DMP models. Next, we describe how we discipline parameters directly related to our countercyclical congestion mechanism—the aggregate production function and the associated worker type evolution.

4.1 Standard Parameters

We set the discount factor to $\beta = 0.99$, which yields an annual real interest rate of about 4%. The matching function takes on the Cobb-Douglas form, $M(u, v) = \bar{m}u^\mu v^{1-\mu}$, where we follow Shimer (2005) and set $\mu = 0.72$. Matching efficiency \bar{m} is set such that the model matches the average US empirical quarterly job finding rate of 0.57. We impose the Hosios condition and set the bargaining power of workers equal to the elasticity of the matching function, $\phi = \mu$.²² Finally, the vacancy posting cost κ is set such that labor market tightness is normalized to $\theta = 1$ in steady state.

The flow value of unemployment b is set such that the replacement rate (relative to the average wage) is 40%, as in Shimer (2005), which gives $b = 0.39$. Hence, our parameterization is not based on a low (fundamental) surplus, which determines the amplification of productivity shocks in the standard model (see e.g., Ljungqvist and Sargent, 2017). Instead, amplification from countercyclical congestion works through a more volatile allocative productivity of new hires.

4.2 Aggregate Shocks

Aggregate productivity z and job separation rate δ follow AR(1) processes in logs,

$$\ln(z_{t+1}) = (1 - \rho_z) \ln(\bar{z}) + \rho_z \ln(z_t) + \sigma_z \varepsilon_{t+1}^z \quad (21)$$

$$\ln(\delta_{t+1}) = (1 - \rho_\delta) \ln(\bar{\delta}) + \rho_\delta \ln(\delta_t) + \sigma_\delta \varepsilon_{t+1}^\delta, \quad (22)$$

²¹To be consistent with our discrete time model, transition rates are not adjusted for time aggregation bias. Appendix A.2 reports how our measured flows compare to adjusted flows and that our data are essentially the same as those used by Shimer (2012).

²²The Hosios condition holds exactly when $\sigma = 1$; with congestion ($\sigma < 1$), surplus may also depend on labor market tightness through marginal products out of steady state. For a special case of the generalized Hosios condition see Mangin and Julien (2020).

Table 1: Parameter Values

Parameter	Value		Target/Source	Data	Model
	Congestion	No congestion			
β	0.99		Annual interest rate	0.04	0.04
μ	0.72		Shimer (2005)	0.72	0.72
\bar{m}	0.57		Job finding probability	0.57	0.57
η	0.72		Hosios condition	0.72	0.72
b	0.39		Avg. replacement rate	0.40	0.40
κ	0.21		Normalization $\theta = 1$	-	1.00
\bar{z}	1		Normalization	-	1.00
σ_z	0.008		St. dev. of ALP	0.010	0.010
$\underline{\rho}_z$	0.956		Persistence of ALP	0.74	0.69
$\bar{\delta}$	0.037		Unemployment rate	0.063	0.063
σ_δ	0.107	0.080	St. dev. of UE/E	0.067	0.067
ρ_δ	0.709	0.670	Persistence of UE/E	0.84	0.74
$\rho_{\delta,z}$	-0.505	-0.560	corr(ALP, δ)	-0.41	-0.41
σ	0.241	1	Impulse response of θ to δ , see Figure 3		
α_k	see Appendix E		$p_k = 1$ for all k in steady state		

Note: The table reports parameter values for both the baseline "congestion" model ($\sigma = 0.241$) and for the standard "no-congestion" model ($\sigma = 1$). The calibration targets are the same across the two parameterizations, with the exception of the IRF of θ to δ , which is not targeted in the no-congestion model.

where \bar{z} and $\bar{\delta}$ are the means, $\rho_z, \rho_\delta \in (0, 1)$ are the persistence parameters, $\varepsilon^z, \varepsilon^\delta \sim N(0, 1)$ are standard-normal innovations to the productivity and separation processes, and $\sigma_z, \sigma_\delta > 0$ are their respective standard deviations.

Aggregate Productivity Shocks. We normalize average TFP, \bar{z} , to one. In order to pin down the persistence and volatility parameters, we target the observed autocorrelation and standard deviation of average labor productivity—ALP (real output per worker in the non-farm business sector). With congestion, ALP in our model is an endogenous object and therefore these parameters guiding the TFP process are estimated jointly with the remaining ones, as described in Section 4.3.

Separation Shocks. The average separation rate, $\bar{\delta}$, is set such that the model matches an average unemployment rate of 6.3% for our sample period of 1976–2019. Given the prominence of UE flows for our novel congestion channel, we target the volatility and persistence of the share of UE flows in employment.

Finally, we let the correlation between ε_z and ε_δ , $\rho_{\delta,z}$ be such that the model matches the correlation between average labor productivity and the separation rate observed in the data. We parameterize the aggregate shock processes jointly with the congestion parameter σ , which we describe below, as the behavior of UE flows is an equilibrium outcome.

Separation Shocks: Alternative Parameterization. We consider an alternative parameterization of separation shocks in which we directly target the empirical δ (EU separation rate) process and account for the impact of non-participation flows. In particular, we add realistically cyclical and exogenous flows between non-participation and unemployment. While this version abandons the convenient structure of our baseline two-state labor market model, it allows us to isolate the variation in UE flows from non-participation flows which are otherwise picked up by separation shocks. Importantly, the next section shows that this version of the model features realistic business cycle variation of the separation rate while, at the same time, leaving our amplification results intact. We discuss this alternative further in Section 5.1 and provide additional details in Appendix I.

4.3 Parameters Related to Congestion in Hiring

We now describe our calibration strategy for model parameters directly related to our congestion mechanism. These include the specification of worker types and their evolution, as well as the parameter governing the elasticity of substitution between worker types.

Number of Worker Types, K . We set a maximum of $K = 160$ steps, i.e., 40 years, after which employed workers remain in the highest rung of the type ladder. Of course, hardly any worker reaches this type given the separation rate.

Since employed workers climb the type ladder by one rung every period, the absorbing (maximum) step K guides how long cohorts preserve their “abundance.” In Appendix G, we explore

the role of K and show that our model preserves its quantitative performance for a wide range of K values. This holds true even without recalibrating the key parameter governing the substitutability of worker types, σ .

Worker Type Evolution: Full Downgrading to $k = 1$. In our baseline specification—without loss of quantitative generality—we assume full type downgrading in unemployment, i.e., $k_u(k) = k - 1$. This process is consistent with the interpretation of worker heterogeneity as reflecting the accumulation and decumulation of skills as in turbulence models (see e.g., Ljungqvist and Sargent, 1998, 2004; den Haan, Haefke, and Ramey, 2005, who permit gradual skill decline, although in these models all worker skill types are perfect substitutes in production).

In Section 4.4, we show robustness to an alternative downgrading specification, in which a certain fraction of workers does not incur any downgrading at all. We show analytically that this model variant, once recalibrated to match the same targets, is *isomorphic* to our baseline specification.

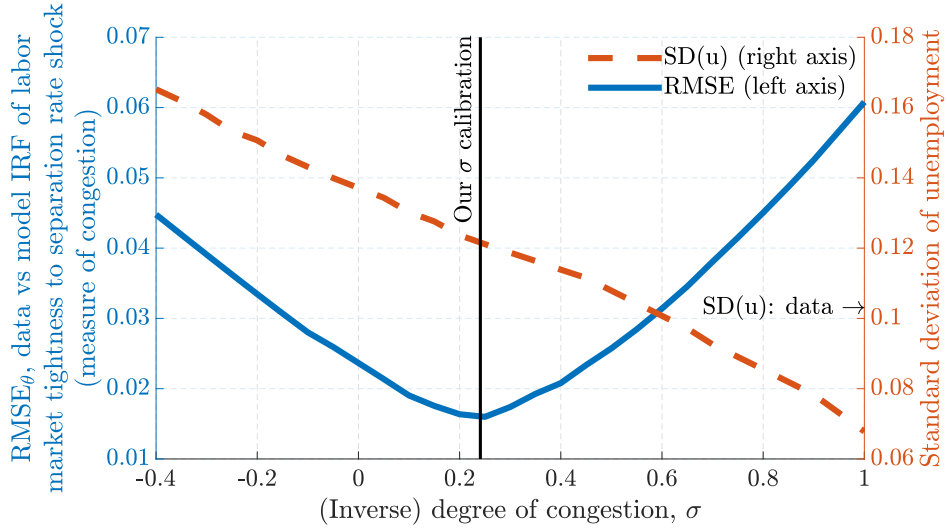
Type-Specific Weights in Production. Each worker type k enters the aggregate production function with its own, type-specific, weight α_k . We use these parameters to ensure comparability of steady-states across different model variants (e.g., when considering different values of σ). In particular, we require that *steady-state* surpluses of all worker types are identical by ensuring $p_k = 1$ for all k . Appendix E provides further details. Combined with our previous choices, this parameterization naturally nests the large-surplus calibration of Shimer (2005) as a special case when $\sigma = 1$.

Disciplining Congestion Parameter σ : IRF of Labor Market Tightness to Separation Shocks. Congestion is guided by the parameter that governs the elasticity of substitution between worker types, σ , which determines the degree of diminishing returns to specific worker types. We parameterize σ (jointly with other model parameters, in particular the α_k -skill weights to maintain the unit MPL for each cohort in steady state) by having the model match the impulse response of labor market tightness to a separation shock, estimated using the same VAR as in Section 2 on simulated data from the model. To do so, we minimize the root mean squared error (RMSE) between the empirical and model impulse responses. Figure 6 plots, as the blue solid line, how this RMSE varies with the congestion parameter σ . We obtain the best fit at $\sigma = 0.241$.²³ The figure also shows the amplification generated by the model, by means of plotting unemployment volatility on a secondary axis, which we return to in the next subsection.

Figure 3 plots the IRF of labor market tightness to a separation shock in the calibrated model, with $\sigma = 0.241$, along with the empirical IRF. The model matches the empirical pattern well.

²³Negative values of σ imply an elasticity of substitution $1/(1 - \sigma) < 1$, i.e., worker types (or equivalently intermediate inputs) are gross complements in final good production. For estimates of *long-run* elasticities of substitution between education and experiences groups in the labor market, see, e.g., Katz and Murphy (1992); Jeong, Kim, and Manovskii (2015); Bils, Kaymak, and Wu (2020). These estimates do not provide a direct comparison to our value of σ as the skill types do not map into our notion of worker types and as they do not use short-run variation.

Figure 6: Calibrating Congestion Parameter σ



Notes: For various values of congestion parameter σ , the figure plots the root mean squared error between the data and model impulse responses of market tightness to a job separation shock (left axis) and the standard deviation of the unemployment rate (right axis). We highlight our baseline calibration with the vertical line.

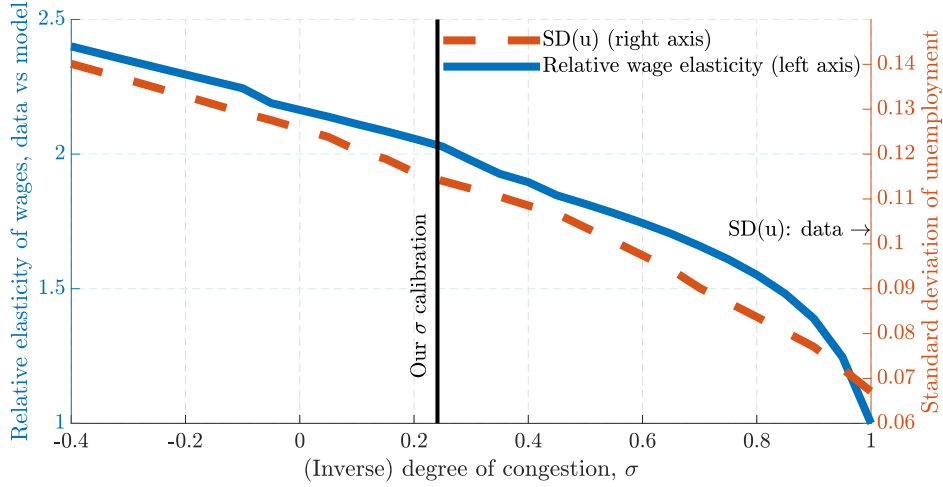
Besides capturing the large negative impact response, the model also generates the observed persistent, hump-shaped dynamics of labor market tightness. The figure further plots the IRF of the standard model without congestion ($\sigma = 1$). That IRF is essentially flat at zero, quantitatively confirming that the equilibrium DMP model exhibits patterns approximated well by the simple no-congestion benchmark discussed in Section 2. Crucially, the inability of the no-congestion model to match the IRF is not a matter of calibration. In Appendix F, we show analytically and by presenting simulated moments that even an alternative calibration with a low surplus in the spirit of Hagedorn and Manovskii (2008) cannot do better; specifically, the model continues to produce the counterfactually flat IRF to separation shocks.²⁴

Validation and Alternative Calibration of σ : The Excess Cyclicity of New-Hire Wages. An alternative calibration strategy is to directly discipline the parameter guiding the congestion mechanism, and in turn the relative productivities of new hires compared to the average worker. One possibility of doing so is by matching the relative wage cyclicality of newly hired and average workers.

Figure 7 follows the structure of Figure 6, but now plots, with the blue solid line, the wage cyclicity of new hires relative to those of all workers for the same range of σ values. In particular, for each value of σ , we simulate the model and construct the semi-elasticity of log wages with respect to the unemployment rate, separately for new hires and for the average worker. In the standard model without congestion, where $\sigma = 1$, all hiring cohorts are perfect substitutes, and

²⁴Additionally, Appendix Figure A11 depicts the impulse response of unemployment to a separation shock. Our congestion model exhibits a much stronger response of unemployment than the standard model without congestion, exactly because labor market tightness falls, which pushes down the job finding rate.

Figure 7: Relative Procyclicality of New-Hire Wages Compared to Average Wages



Notes: For various values of congestion parameter σ , the figure plots the relative wage cyclicality of new hires and the average worker (the ratio of the semi-elasticity of wages to the unemployment rate) on the left axis, and the standard deviation of the unemployment rate on the right axis. We highlight our baseline calibration with the vertical line.

hence have homogeneous productivities and wages. The semi-elasticity ratio is therefore one, depicted as the rightmost value of σ .

When $\sigma < 1$, new hires' wages are relatively more procyclical because UE flows increase in recessions, lowering relative productivity in new jobs. The bargained wages reflect this productivity differential. For our preferred value $\sigma = 0.241$, the model exhibits an excess procyclicality of new hires' wages of around two. Reassuringly, this value falls into the range of relative wage cyclicality observed in the US micro data, as reported by the canonical meta-analysis in Pissarides (2009) (Table II therein).²⁵ Importantly, as the red dashed line and secondary y-axis reiterate, this relative wage semi-elasticity is with respect to a realistic value of unemployment rate volatility.

While these results are encouraging, we choose not to pursue this line of parameterization as our baseline strategy because we believe it faces several limitations. First, the degree to which wages reveal idiosyncratic productivity depends on the bargaining power of workers, which we set to a relatively high value following the macro literature (compared to micro-evidence on, e.g., rent sharing elasticities, see, e.g., Jäger, Schoefer, Young, and Zweimüller, 2020). Similarly, in logs, the wage cyclicity depends on the level of the surplus $p - b$, where we assume homogeneous, acyclical outside options (for an empirical critique, see Chodorow-Reich and Karabarbounis, 2016). Second, our model does not feature wage rigidity, and thereby loads all wage cyclicity into the channel of differential productivities. Wage rigidity for incumbents only (e.g., Beaudry and DiNardo, 1991; Shimer, 2004; Schoefer, 2021) would lead us to underestimate σ (overestimate congestion); wage rigidity for all workers Hall (2005b) would mean we overestimate σ (underestimate congestion).

²⁵Recall that our model is calibrated such that all worker types have identical wages in steady state ($p_k = 1$ and hence $w_k = w$ for all k), so our model-based wages are by construction not subject to composition effects, and hence correspond to the estimates in Pissarides (2009), which are composition-adjusted for worker quality (see, e.g., Bils, 1985; Haefke, Sonntag, and van Rens, 2013).

Third, while we here study flow entry wages (and assume period-by-period bargaining in our model), it is the present value of new hires' wages, potentially spread out in various ways over time, that is allocative for hiring and would robustly reflect productivity differentials (Shimer, 2004; Kudlyak, 2014), which we address in Section 6.2. Finally, to the extent that diminishing returns in the aggregate production function manifest themselves in the expansion of low-productivity and low-wage jobs or firms in recessions, our mechanism may evade the ongoing debate about the role of job and firm composition in new hires' wage fluctuations (Pissarides, 2009; Gertler, Huckfeldt, and Trigari, 2020; Hazell and Taska, 2020; Grigsby, Hurst, and Yildirmaz, forthcoming).

Cross-Check: Comparison with Existing Estimates on Substitution Between Labor Types. To our knowledge, there is no independent and direct evidence on the value of σ for the labor types we are considering (new hires and incumbent workers). However, it is still useful to benchmark the level of σ against related estimates in the literature, which studies parameters that guide the elasticity of substitution between labor types different than the ones we are considering, namely educational groups and workers of different age and experience levels. Appendix Table A3 presents these parameter estimates that would correspond, for new hires vs. incumbent workers, to our congestion parameter σ . All estimates draw on a similar CES aggregate production function.²⁶ Our estimate of $\sigma = 0.241$ lies well within the range of estimates in the literature. While the labor types and estimation strategies in those papers differ from our context, we cautiously conclude that the literature predominantly estimates different types of workers to be gross substitutes—as we do for our types—and at levels broadly similar to ours.

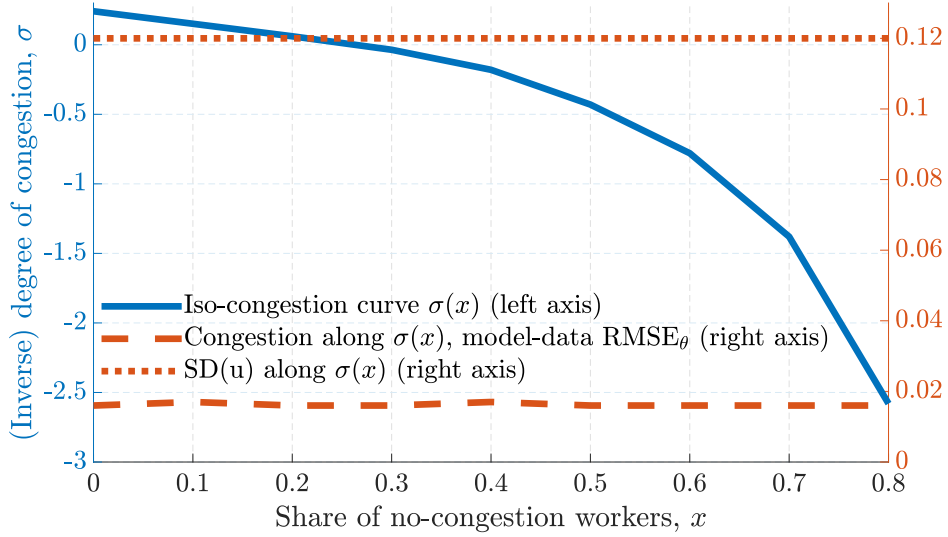
4.4 Robustness to Alternative Congestion Mechanisms

Here we show that alternative model structures yield similar properties to our baseline model, as long as the parameter guiding congestion is recalibrated to match the empirical IRF of labor market tightness to a separation shock.

A Model Featuring Both Congestion and Non-Congestion Hires. Our baseline model features a parsimonious skill process: job loss resets worker types to $k = 1$. In reality, a fair share of the unemployed may enter reemployment in their original type, e.g., due to not losing their skill, being hired directly into higher-level positions, or being recalled. Such departures may seem to reduce amplification. However, for such model variants to still match the empirical degree of congestion, our calibration strategy simply would estimate a lower σ parameter, and ultimately exhibit the same degree of congestion.

²⁶The literature typically starts by positing an aggregate production function, assumes a perfectly competitive labor market and derives an estimating equation that relates the wages of different types of workers to their relative quantities in production. The coefficient on the relative quantities of worker types is a function of the parameter that governs the elasticity of substitution between these types. The literature then estimates the derived wage-premium equation and backs out the implied parameter that we report in the last column of Appendix Table A3.

Figure 8: Robustness to Alternative Specifications of Skill Process



Notes: The figure plots recalibrated values of σ for different shares of no-congestion hires, x , the “iso-congestion” curve $\sigma(x)$. It also plots the RMSE between the empirical and model-implied IRF of labor market tightness to separation shocks, and the standard deviation of unemployment for the recalibrated models to highlight that congestion and amplification properties of the model stay the same as long as σ is recalibrated to match the market-tightness impulse response target.

To demonstrate robustness, in Appendix J, we elaborate on an alternative to the type evolution we assume in the baseline model. In particular, fraction x of “no-congestion hires” replicate the skill structure prevailing at the point of hiring; fraction $1 - x$ of “congestion hires” fully downgrade to $k = 1$. Isomorphically, the no-congestion workers operate in a separate linear production function.²⁷

Figure 8 shows two model properties for different values of x . $x = 0$ gives our baseline model, where all workers entail congestion. Importantly, each x -model is reparameterized to match all the calibration targets, including the RMSE target (red dashed line). To achieve this fit, a higher x model simply requires a lower and lower σ —traced out by the “iso-congestion” $\sigma(x)$ curve (blue solid line), derived analytically in Appendix J.²⁸

Importantly, the dotted red line shows that the standard deviation of unemployment is invariant in x along the iso-congestion curve for σ —so such alternative specifications of the worker-type process are *isomorphic* to our baseline specification in which all workers fall to $k = 1$ upon job loss.

²⁷Here, the α_k -skill weights are recalibrated to yield homogeneous productivities in steady state. In this second interpretation, the final good is produced as a convex combination of congestion (CRS-CES) and no-congestion (linear) production functions, $Y = z[(1 - x)(\sum_{k=1}^K \alpha_k^c (n_k^c)^\sigma)^{1/\sigma} + x(\sum_{k=1}^K \alpha_k^{nc} n_k^{nc})]$, where subscripts c and nc stand for the congestion and no-congestion sectors.

²⁸There, we consider a simple analytical expression for the elasticity of the marginal product of an *average* new hire \bar{p}_1 as a function of cohort size n_1 : $\varepsilon_{\bar{p}_1, n_1} = (\sigma - 1)(1 - n_1/N)(1 - x)$. The iso-congestion curve for a desired degree of congestion $\bar{\varepsilon}$ as a function of no-congestion worker share x is given by $\sigma(x, \bar{\varepsilon}) = 1 + \frac{\bar{\varepsilon}}{(1-x)(1-n_1/N)}$. This analytical curve turns out to be essentially identical to the blue line.

Congestion Through Convex Hiring Costs. In Appendix K, we present a structurally more divergent model, in which congestion operates through a convex cost in gross UE hires, rather than through the production function.²⁹ All workers are perfect substitutes and homogeneous. Again, once this model variant is calibrated to exhibit realistic congestion in hiring, it too generates similar cyclical patterns of key labor market variables. The intuition is that the countercyclical employment share of UE hires increases the hiring cost during recessions. This property stands in contrast to the procyclicality of recruitment costs in the standard no-congestion model, which lead to dampening rather than amplification (as explained in, e.g., Shimer, 2010). However, unlike our baseline framework with productivity-based congestion, the model with convex hiring costs does not generate more cyclical wages of new hires compared to average wages, nor can it speak to the additional applications we study in Section 6.

5 Quantitative Performance over the Business Cycle

We now study the quantitative implications of countercyclical congestion for labor market fluctuations. We begin by describing basic business cycle statistics for our baseline model and—for comparison—several alternatives. Thereafter, we focus on specific business cycle properties, including a detailed description of the sources of amplification our baseline model provides.

5.1 Business Cycle Statistics

The main results are summarized in Table 2. For compactness, the table only reports standard deviations, autocorrelations and correlations with the unemployment rate; we relegate the full correlation matrices to Appendix H.

Empirical Benchmark. As the empirical benchmark, Table 2 Panel A provides an overview of business cycle statistics for quarterly US data.

Theoretical Benchmark: No-Congestion Models. Panels B and C of Table 2 report on the cyclical behavior of the no-congestion model, which is isomorphic to the standard DMP model calibrated as in Shimer (2005).³⁰ Panel B reports on the standard DMP model without separation shocks. As is well understood, this model does not provide enough amplification of labor market tightness (just 7% of that in the data), such that the job finding rate and unemployment are counterfactually smooth. Incorporating separation shocks into the no-congestion model helps along several dimensions, as Panel C shows. Most notably, the correlation of UE flows and unemployment becomes positive and close to that in the data (0.74). With separation rate shocks, the volatility of unemployment increases, but insufficiently so, with the additional volatility largely driven by the

²⁹See, e.g., Fujita and Ramey (2007); Coles and Moghaddasi Kelishomi (2018); Mercan and Schoefer (2020) for models that modify the free-entry condition along those lines.

³⁰The exception is the HP-filter smoothing parameter, which we discuss in Appendix A.4.

Table 2: Business Cycle Properties: Data and Model Variants

	ALP	f	δ	u	v	θ	UE/E	p_1
<i>Panel A: Data</i>								
Standard deviation	0.010	0.053	0.067	0.103	0.126	0.229	0.067	
Autocorrelation	0.746	0.871	0.773	0.934	0.926	0.936	0.836	
Correlation with u	-0.112	-0.931	0.848	1	-0.934	-0.980	0.833	
<i>Panel B: No-Congestion Model Without Separation Shocks</i>								
Standard deviation	0.010	0.004	0	0.003	0.013	0.015	0.003	0.010
Autocorrelation	0.704	0.704	0	0.843	0.592	0.704	0.306	0.704
Correlation with u	-0.643	-0.643	0	1	-0.481	-0.643	-0.272	-0.643
<i>Panel C: No-Congestion Model With Separation Shocks</i>								
Standard deviation	0.010	0.005	0.088	0.068	0.058	0.017	0.067	0.010
Autocorrelation	0.688	0.647	0.499	0.736	0.751	0.647	0.740	0.688
Correlation with u	-0.508	-0.665	0.916	1	0.974	-0.665	0.739	-0.508
<i>Panel D: Congestion Model—Baseline (Matching UE/E)</i>								
Standard deviation	0.010	0.059	0.122	0.121	0.102	0.207	0.067	0.055
Autocorrelation	0.688	0.897	0.530	0.836	0.857	0.897	0.742	0.771
Correlation with u	-0.463	-0.924	0.743	1	-0.716	-0.940	0.865	-0.862
<i>Panel E: Congestion Model—Robustness (Matching EU & Participation)</i>								
Standard deviation	0.010	0.054	0.067	0.099	0.099	0.189	0.052	0.051
Autocorrelation	0.701	0.901	0.544	0.850	0.889	0.902	0.767	0.781
Correlation with u	-0.337	-0.941	0.693	1	-0.819	-0.954	0.890	-0.882

Notes: ALP , f , δ , u , θ , UE/E and p_1 indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, share of new hires in employment and the marginal product of labor of new hires (which is identical to ALP in the no-congestion models). Panel A reports values from the data; Panels B and C report these values for the no-congestion model without and with shocks to the separation rate. Panel D reports the results for our baseline congestion model (which matches the employment share of UE hires). Panel E shows robustness to a congestion model that instead matches the EU separation rate fluctuations (and includes a participation margin). All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600. Appendix H reports the full correlation matrices.

inflows.³¹ Moreover, adding separation shocks into the no-congestion model comes at the cost of a wrongly signed Beveridge curve (as discussed in the next section).

The Baseline Congestion Model. Panel D reports moments from our congestion model. The model closely replicates the business cycle properties of the key empirical variables, both with regards to volatility and cyclical. Specifically, we have a nearly perfect fit of the standard deviation of unemployment, labor market tightness and, accordingly, the job finding rate. Hence, countercyclical congestion can be viewed as a solution to the inability of the standard DMP model to generate realistic labor market fluctuations (Shimer, 2005).

³¹In Appendix F, we show analytically that the no-congestion model's elasticity of labor market tightness to the separation rate is small in a broad class of model parameterizations.

The Degree of Congestion (σ) and Unemployment Fluctuations. Figure 6 complements the table, visualizing how congestion leads to amplification. The figure additionally plots, with a red dashed line, the volatility of unemployment for different values of σ (while recalibrating all other parameters to match the remaining targets), so that we surgically isolate the role of congestion.³² Intuitively, as separations increase in a recession, unemployment rises. As a result, UE flows rise, which, as long as $\sigma < 1$, pushes down the productivity of new hires, compared to average labor productivity (last column of Table 2 Panel D). This productivity drop further reduces hiring incentives, keeping unemployment elevated—a mechanism mediated by congestion, i.e., the level of σ .

Robustness: Matching EU Separations and Non-Participation Flows. The congestion model does estimate higher volatility and lower auto-correlation of the separation shock δ compared to the data. This artifact emerges because we choose the convenient structure of a two-state labor market model, ignoring flows into and out of non-participation (see, e.g., Elsby, Hobijn, and Şahin, 2015a). Therefore, making our model match UE flows in the data (one of our baseline calibration targets) results in separation shocks picking up the ignored variation in the participation margin. When we instead account for such flows, the EU separation process turns fully realistic while—importantly—leaving our amplification results intact (see Appendix I and Panel E in Table 2).

Further Robustness: Small Surplus. Appendix F studies the no-congestion model under the Hagedorn and Manovskii (2008) calibration, i.e., featuring a small match surplus in steady state (high b relative to productivity), which permits productivity shocks to have a large effect on hiring and generate realistic labor market volatility (Ljungqvist and Sargent, 2017). We have additionally experimented with a model featuring decreasing returns in aggregate employment, similar to Michaillat (2012). Both of these two model variants, however, would predict essentially no effect of separation rate shocks on labor market tightness—as in the standard DMP model—for lack of a congestion in hiring, i.e. again generating a wrongly sloped Beveridge curve.

5.2 Beveridge Curves

We now study the Beveridge curve, i.e., the relationship between vacancies and unemployment (see Elsby, Michaels, and Ratner, 2015b, for a review). In fact, the Beveridge curve highlights the core difference between congestion and no-congestion models.

Figure 9 plots the Beveridge curves of the congestion model ($\sigma = 0.241$), the data, as well as the standard, no-congestion ($\sigma = 1$) model. In the data, the Beveridge curve is negatively sloped, with a correlation of -0.934 and standard deviations of 0.126 and 0.103 for vacancies and unemployment respectively, as reported in Table 2 Panel A.

³²Concretely, we maintain the same, high fundamental match surplus for each model. We do so by recalibrating the productivity weights α_k to generate the common unit productivity in steady state for all types, as in Section 4.

The no-congestion model with separation rate shocks features a counterfactually positive slope (a 0.974 correlation): as unemployment increases, vacancies *rise*. In the model, fluctuations arise from two shocks, namely shocks to TFP and the separation rate. TFP shocks on their own would lead to a negative slope, but these hiring-induced fluctuations are small due to insufficient amplification (Shimer, 2005). Instead, separations drive unemployment fluctuations here; but the no-congestion model exhibits a counterfactually flat IRF of labor market tightness to a separation rate shock, as described in Sections 2 and 4.3. On net, separation shocks dominate in this model, tilting the Beveridge curve into the wrong direction (see also Shimer, 2005).

By contrast, the congestion model closely matches the empirical negatively sloped Beveridge curve (a -0.716 correlation for the baseline congestion model, and -0.819 for the alternative specification matching separation rate fluctuations). This success is at the heart of how congestion affects the overall dynamics of the labor market: in our model, separation shocks lead to large and persistent increases in unemployment. They do so by incipiently raising UE flows, i.e. gross flows back into employment, exactly as in the no-congestion model. But in the congestion model, precisely this process of expanding gross flows diminishes the returns to further hiring, permitting the model to rationalize elevated unemployment. Overall, our model still falls short of perfectly matching the empirical slope of the Beveridge curve, indicating that further departures from free entry may be needed.³³

5.3 Sources of Amplification: Productivity and Cohort Dynamics

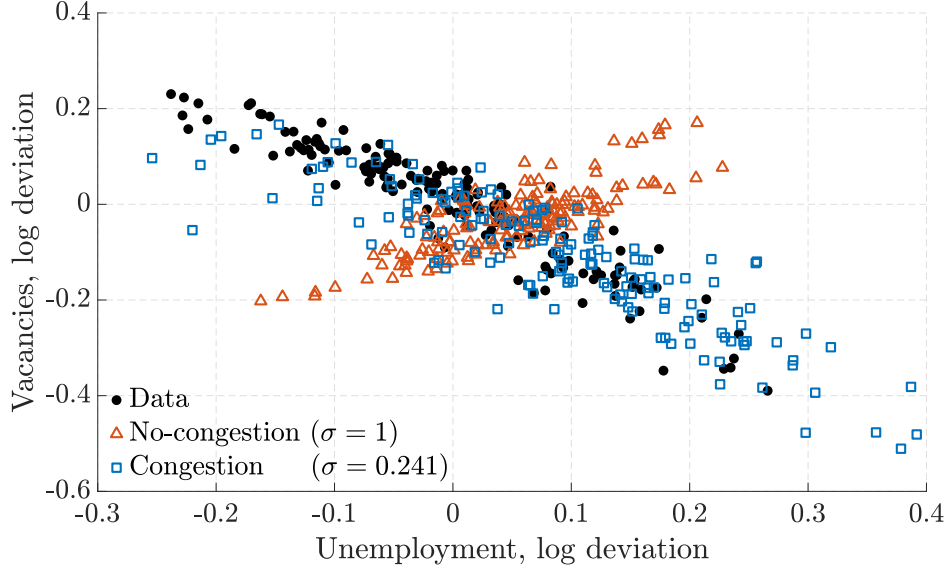
The key to understanding amplification is the behavior of the match surplus for new hires. Using Equation (16) and imposing the assumption that $k_u(k) = k - 1$ (i.e., full type downgrade), we can simplify the surplus expression for any worker type k as

$$S_{k,t} = \underbrace{p_{k,t}}_{\text{Current productivity}} - b + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) S_{k+1,t+1} \right]}_{\text{Continuation value at } k+1} - \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) f(\theta_t) \phi S_{1,t+1} \right]}_{\text{Worker's outside option: finding new job next period at } k=1}. \quad (23)$$

In comparison to the no-congestion model, amplification in surplus fluctuations stems from three sources. First, the *flow productivity channel* works through a more volatile and procyclical productivity of new hires, compared to the standard measure of average labor productivity. Second, two dynamic effects emerge through cohort effects: the *present value channel* through the continuation value of employed workers, and the *outside option channel*. We rearrange the surplus expression in Equation (23) to explicitly highlight these three amplification channels, now specifically focusing

³³Coles and Moghaddasi Kelishomi (2018) too obtain a correctly sloped Beveridge curve despite time-varying separations. Their mechanism works through the unemployed depleting the stock of vacancies due to inelastic free entry (vacancy creation). See also Elsby, Michaels, and Ratner (2015b) for a discussion. Part of the remaining gap indicated in Table 2 Panel D however reflects the fact that our calibration strategy selects too volatile a separation rate shock process to match the share of UE flows in employment; Panel E reports the version of the model matching the separation rate process directly, where the slope of the Beveridge curve is even more realistic.

Figure 9: Beveridge Curve: Data and Models



Notes: The figure is a scatter plot of quarterly time series of vacancies (normalized by the labor force), based on the vacancy time series by (the Composite Help Wanted Index by Barnichon, 2010), against the unemployment rate in the US data from 1976 to 2016, and the simulated time series from the no-congestion and congestion models. All variables are logged and HP-filtered using a smoothing parameter of 1,600.

on the surplus of new hires $k = 1$:

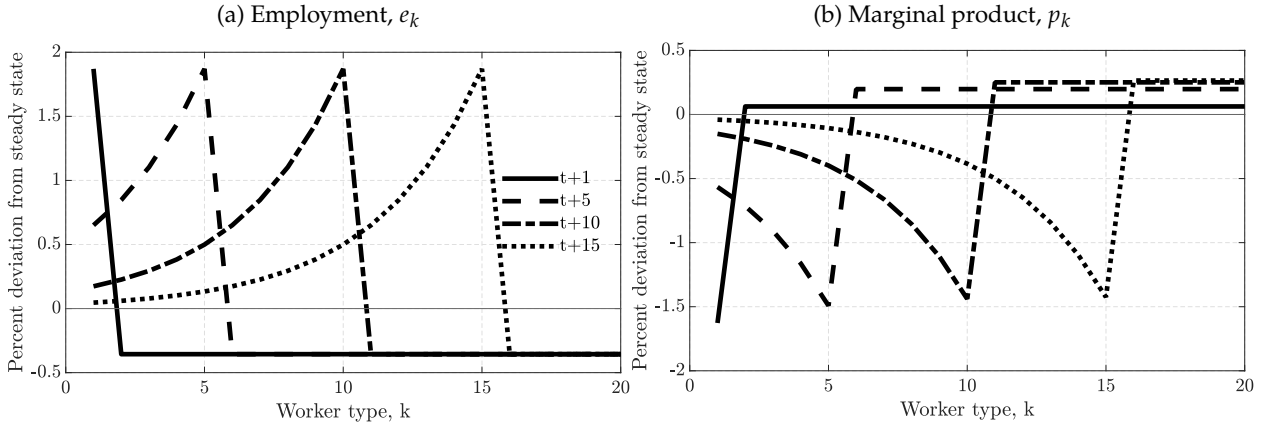
$$\begin{aligned}
 S_{1,t} = & \underbrace{z_t - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t^s)) S_{t+1}^s \right]}_{(i) \text{ No-congestion model surplus}} + \underbrace{S_t^* - S_t^s}_{(ii) \text{ Flow productivity channel}} \\
 & + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t)) \phi (S_{2,t+1} - S_{t+1}^*) \right]}_{(iii) \text{ Present value channel}} + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) f(\theta_t) \phi (S_{2,t+1} - S_{1,t+1}) \right]}_{(iv) \text{ Outside option channel}}, \quad (24)
 \end{aligned}$$

where $S_t^s = z_t - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t^s)) \phi S_{t+1}^s \right]$ is the surplus in the standard model without congestion and homogeneous workers, and θ^s is the associated labor market tightness.³⁴ $S_t^* = p_{1,t} - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t)) \phi S_{t+1}^* \right]$ is the match surplus in which flow productivity is (counterfactually) always equal to that of new hires, $p_{1,t}$. We now investigate the three new sources of amplification (ii)–(iv) in detail.

Flow Productivity Channel. As foreshadowed in Figure 5, Table 2 shows that countercyclical congestion dramatically amplifies the productivity of new hires, which is around five times as volatile as—and masked by the smoothness of—average productivity. It is also more procyclical, with a correlation with unemployment of -0.862 , compared to -0.463 for average productivity.

³⁴That is, for this standard surplus term, we use the counterfactually smooth job finding rate generated by the standard model to construct the standard surplus. All other terms use the same job finding rate generated by the congestion model.

Figure 10: Impulse Responses by Type k : Employment and Productivity



Notes: The figure plots impulse responses across types of employment and marginal productivities by worker type (only first 20 types are shown) to a perfectly transitory separation shock. Each line represents the cross-sectional response in a particular point in time. All variables are expressed in percent deviations from their respective steady states.

Intuitively, UE flows rise in recessions, so that new hires become abundant, which lowers their marginal product.

Cohort Effects: Present Value Channel. New hires in recessions are not just congested in the first period. Instead, persistent *cohort effects* arise, as new hires stick with their initial cohort size as they move up the rungs of the type ladder together.

Figure 10 visualizes these cohort effects by depicting the impulse response, to a perfectly transitory separation shock, of employment and productivity of different worker types k . Each line represents the deviation from steady state for a particular period. For instance, the solid line shows the response for workers newly hired in the period, i.e., $t + 1$. Because of the inflow of new hires, employment of the lowest type, $k = 1$, expands (Panel (a)). This abundance pushes down their productivity (Panel (b)). These spikes persist throughout the affected cohort's tenure. For example, the workers that survive from the abundant cohort of newly hired ($k = 1$) workers in period $t + 1$ become the—still abundant—cohort of $k = 2$ type workers in period $t + 2$ and so on.³⁵ As a result of these persistent cohort effects, the expected present value of productivity of newly hired workers—which is allocative for hiring—essentially inherits the excess volatility of flow productivity, and is indeed almost five times as volatile as in the standard model without congestion. This is the key driver of variation in term (iii) in Equation (24), since $S_{2,t+1}$ incorporates cohort effects, while S_{t+1}^* does not.

Cohort Effects: The Outside Option Channel. Cohort effects generate a second dynamic impact on surplus fluctuations, operating through workers' outside options in bargaining. A new hire,

³⁵The slight recovery in their productivity is solely due to the recovery in total employment, as separations slightly shrink all other types upon impact, namely incumbents.

Table 3: Volatility of Labor Market Tightness and Sources of Amplification

		Standard deviation	Contribution to total
No-congestion model	(i)	0.019	0.049
+ Flow productivity channel	(i)+(ii)	0.052	0.162
+ Present value channel	(i)+(ii)+(iii)	0.178	0.851
+ Outside option channel	(i)+(ii)+(iii)+(iv)	0.207	1

Notes: The table reports the standard deviation of (log) labor market tightness in variants of the congestion model. The top row reports values for the standard no-congestion model, the second and third rows incrementally add the productivity and present value channels and the bottom row shows the volatility implied by the baseline congestion model, where all channels are active. The column “contribution to total” shows $\text{cov}(\theta_{\text{base.}}, \theta_{\text{cf.}}) / \text{var}(\theta_{\text{base.}})$, where $\theta_{\text{base.}}$ is labor market tightness in our baseline model, while $\theta_{\text{cf.}}$ is the respective counterfactual labor market tightness.

entering step $k = 1$ at t , has productivity $p_{k=2,t+1}$ at $t + 1$. A new hire at $t + 1$ has an initial productivity of $p_{k=1,t+1}$. At $t + 1$, the differential productivities of these two types depend on their relative abundance at $t + 1$, and similarly for all future periods.

When Nash bargaining, the worker’s outside option is walking away and searching for another job. In the no-congestion model, this outside option moves with the job finding rate, which actually attenuates fluctuations in the surplus value, because $f(\theta)$ falls in recessions, lowering worker’s outside option, thereby expanding surplus.

With congestion and the cohort effects it triggers, the outside option channel reflects additional intertemporal, opportunity-cost considerations. For instance, when congestion is high today but is expected to fall tomorrow, surplus in today’s jobs falls by more than implied by comparing productivity differences.³⁶

Quantifying the Sources of Amplification. To quantify the contributions of the three channels to amplification, we feed in counterfactual surpluses from subsets of the four channels in Equation (24) into the free-entry condition in Equation (20). We report the resulting standard deviations of labor market tightness in Table 3.

The specification with all four channels generates a standard deviation of 0.207, close to the data (see Table 2). In the absence of the outside option channel, the standard deviation remains still high, accounting for 85% of the baseline fluctuations. Therefore, the outside option channel explains only 15% of the fluctuations in labor market tightness. The flow productivity channel, which takes into account the higher volatility of allocative productivity (and that of the implied job finding rate), explains 16% of the variation in labor market tightness. Finally, the no-congestion model accounts for only about 5% of the baseline fluctuations in labor market tightness. Therefore, the strongest effect is through the present value channel, accounting for over 2/3 ($0.851 - 0.162 = 0.689$) of the fluctuations in labor market tightness.

³⁶This mechanism would not be present with wage setting protocols that insulate wages from outside options (Hall and Milgrom, 2008; Jäger, Schoefer, Young, and Zweimüller, 2020).

5.4 Historical Decomposition of Unemployment in the United States

We now study how countercyclical congestion has contributed to empirical unemployment fluctuations in the US since 1976. We do so by feeding into the model an estimated time path of new hires' productivity that would arise only through congestion, i.e., movements in new hires' productivity solely explained by fluctuations in the employment share of UE hires. By contrast, we hold fixed TFP and separation rates. We then construct a *counterfactual* unemployment time series due to this congestion channel alone.

Method. Formally, we use the following equations for counterfactual unemployment, surplus, and labor market tightness that are purely driven by congestion:

$$\begin{aligned}
 u_{t+1}^c &= (1 - f(\theta_t^c))u_t^c + \bar{\delta}(1 - u_t^c) \\
 \kappa &= q(\theta_t^c)\beta\mathbb{E}_t(1 - \bar{\delta})S_{1,t}^c \\
 S_{k,t}^c &= p_{k,t} \cdot \frac{\bar{z}}{z_t} - b + \beta\mathbb{E}_t(1 - \bar{\delta})S_{k+1,t+1}^c - \beta\mathbb{E}_t(1 - \bar{\delta})f(\theta_t^c)\phi S_{1,t+1}^c \text{ for all } k.
 \end{aligned} \tag{25}$$

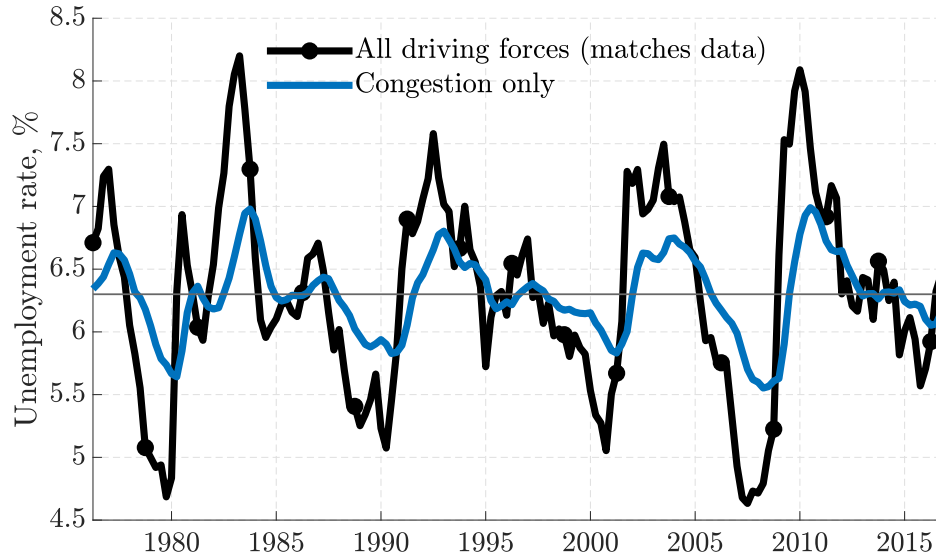
The counterfactual surplus values are based on the congestion model's estimated marginal products $p_{k,t}$, but netting out (i.e., dividing by) aggregate productivity shocks z_t . Hence, the productivity fluctuations that affect surplus are solely due to type-specific congestion, i.e., fluctuations in the employment share of the recently unemployed. Second, we fix the job separation rate at its steady-state value, $\bar{\delta}$. Therefore, u_t^c —"congestion unemployment"—surgically reflects variation due to congestion alone, which we permit to affect the unemployment rate through hiring and the job finding rate.

To obtain historical time series from our congestion benchmark, we use the Kalman filter to estimate the time path of all our model variables (including the marginal products of all worker types $p_{k,t}$) on US time series data for average labor productivity and the share of new hires in employment (logged and HP-filtered with a smoothing parameter of 1,600). Appendix Figure A21 presents both the estimated and empirical time series, and Appendix D contains further details on the estimation procedure. Appendix L provides additional details on the decomposition, and additionally applies the method to TFP-only and separation-only counterfactuals.

The Time Series of Congestion-Driven Unemployment. Figure 11 shows the time path of congestion unemployment in the US, and compares it to overall unemployment (which essentially perfectly tracks the empirical time series, as shown in Appendix Figure A21). First, the autocorrelation coefficient of congestion unemployment is 0.950 relative to 0.905 for overall unemployment, helping generate persistence (Dupraz, Nakamura, and Steinsson, 2019; Hall and Kudlyak, 2020b,a).

Second, congestion is a powerful driver of unemployment fluctuations. The standard deviation of congestion-only unemployment is 0.05, about 40% the level of overall unemployment. Computing the contribution of congestion-only unemployment, we find $\text{cov}(u, u^c)/\text{var}(u) = 0.297$

Figure 11: Historical Decomposition of Unemployment: Actual and Congestion



Notes: The figure plots actual and congestion unemployment (u^c) estimated using data on the cyclical components of average labor productivity and new hires as a share of employment. The counterfactual unemployment time series for u^c is constructed based on the set of Equations (25).

(with a correlation of 0.723). Therefore, countercyclical congestion explains 30 to 40% of observed unemployment fluctuations.

6 Additional Implications of Countercyclical Congestion

Besides providing a new perspective on unemployment fluctuations, countercyclical congestion rationalizes three additional, widely studied, macro patterns: the business-cycle-accounting labor wedge, the countercyclical and persistent earnings losses from job displacement and from graduating in a recession, and the limited long-run sensitivity of labor market variables to labor market policies. To our knowledge, these issues have not been studied *simultaneously* before, and, therefore, the quantitative success of our model in explaining them provides additional external validity to our congestion mechanism.

6.1 Business Cycle Accounting: The Labor Wedge

The Standard Labor Wedge. In a perfectly competitive spot labor market with representative agents, as in RBC models, the household's marginal rate of substitution (MRS) between consumption and labor always equals the marginal product of labor (MPL). In the data, the MRS and the MPL exhibit a strongly cyclical gap, described as a time-varying tax-like *labor wedge* $1 - \tau$ (Chari, Kehoe, and McGrattan, 2007; Shimer, 2009), obtained as a residual—by specifying a utility function and an aggregate production function, and feeding in the empirical time series on consumption C ,

output Y , and employment E —from the following equation:

$$(1 - \tau) \cdot MPL = MRS \left(= \frac{-U_E(C, E)}{U_C(C, E)} \right). \quad (26)$$

This agnostic labor wedge stands for cyclical frictions, mismeasurement or model misspecification. Business cycle accounting (Chari, Kehoe, and McGrattan, 2007) identifies as promising research avenues those refinements that (can be written to) manifest themselves as and replicate the empirical behavior of the labor wedge (and other wedges).

Figure 12 plots the labor wedge time series (red dashed line) calculated using the standard average labor productivity time series (as in Chari, Kehoe, and McGrattan, 2007; Shimer, 2009).³⁷ As is well known, the US data exhibit a volatile and cyclical labor wedge, such that recessions are times when the gap between the MRS and the MPL widens: standard productivity measures fall only slightly, while the MRS falls substantially.

Congestion and the Labor Wedge. To show that the more procyclical marginal product of labor implied by our congestion model offers an explanation for the labor wedge, we first extend our aggregate production function to include capital, \tilde{K} , using a Cobb Douglas specification, with capital share a , and with the labor aggregator mirroring our baseline labor-based CES production function:

$$Y = z\tilde{K}^a \cdot \left(\left[\sum_{l=1}^K \alpha_l n_l^\sigma \right]^{\frac{1}{\sigma}} \right)^{1-a}. \quad (27)$$

Second, to retain comparability to the spot labor market, we consider the productivity of new hires p_1 only. We then reformulate the marginal product of new hires as the standard marginal labor product times a diminishing-returns of new hires term, making clear that this term shows up *exactly* like the labor wedge in Equation (26):

$$MRS = \underbrace{(1-a) \frac{Y}{N}}_{\text{Standard MPL}} \times \overbrace{\frac{\alpha_1 s_1^{\sigma-1}}{\sum_{l=1}^K \alpha_l s_l^\sigma}}^{MPL=p_1} \quad (28)$$

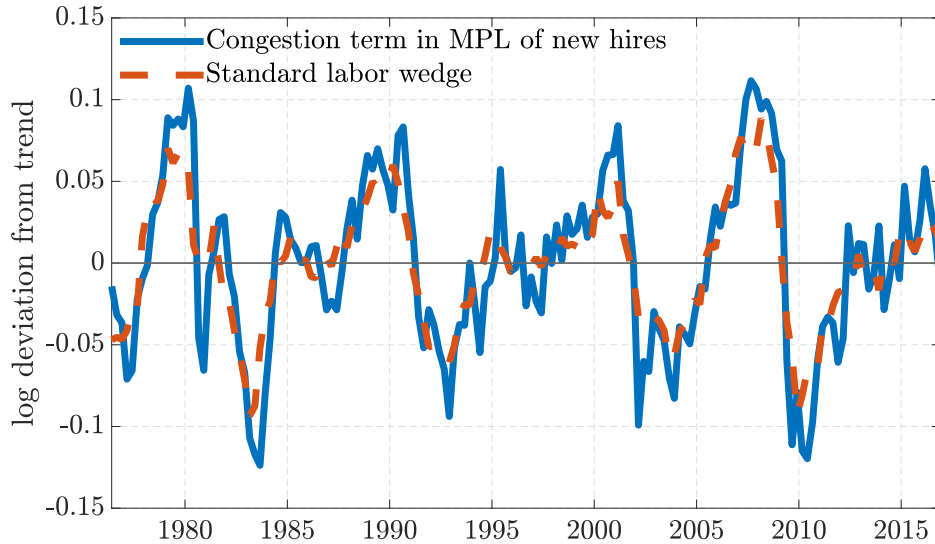
New-hire adjustment term

Figure 12 additionally plots this adjustment term for new hires' productivity (blue solid line). It strikingly closely tracks the standard labor wedge time series (correlation of 0.884).³⁸ The

³⁷Our calculation assumes Cobb Douglas production (as in Chari, Kehoe, and McGrattan, 2007; Shimer, 2009) to construct the MPL as productivity per worker, as our model features only the extensive employment margin. For the household's utility function, we posit separable balanced growth preferences with log consumption utility and a constant Frisch elasticity η of extensive-margin labor supply $U(C, E) = \ln C - \Gamma E^{1+1/\eta}/(1 + 1/\eta)$. We set this elasticity to 0.34, as suggested by Chetty, Guren, Manoli, and Weber (2012).

³⁸We construct the term as the new-hire productivity in Figure 5 Panel (b), described in Footnote 17.

Figure 12: Standard Labor Wedge, and the Productivity of New Hires



Notes: The figure plots the labor wedge implied by the standard productivity measure and the wedge-like productivity adjustment term for new hires in Equation (28). All series are in logs and HP filtered using a smoothing parameter of 1,600.

remaining variation of the labor wedge after subtracting the new-hire term is essentially unrelated to the business cycle: the elasticity of this residual labor wedge variation with respect to the detrended unemployment rate falls to just 0.081 ($R^2 = 0.111$), compared to -0.328 ($R^2 = 0.872$) for the raw labor wedge. That is, the economy with congestion essentially provides a full explanation of the labor wedge.

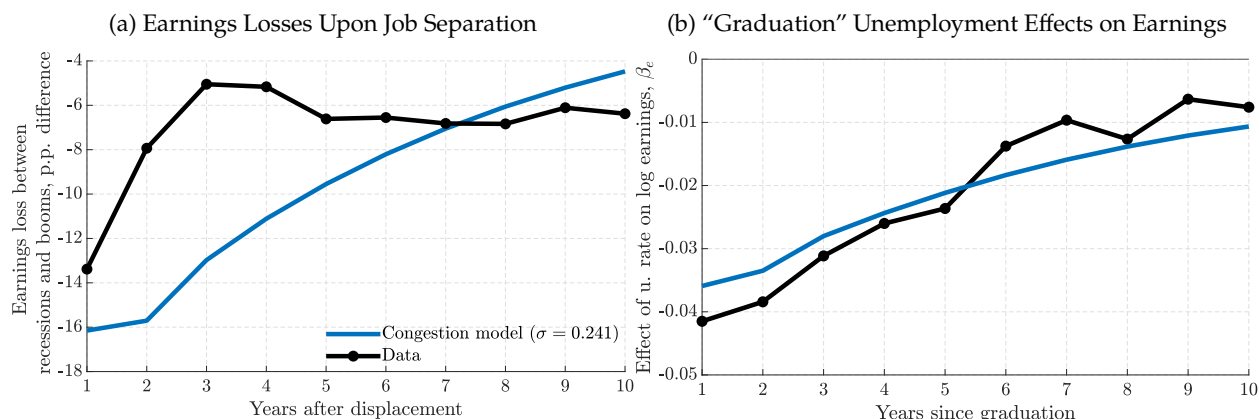
6.2 Countercyclical Earnings Losses From Job Displacement

Our model generates realistically countercyclical earnings losses from job displacement and labor market entry. By additionally highlighting the cohort effects present in our model, this analysis complements that of new hires' flow wages in Section 4.3.

The Cyclicity of Displacement Costs in the Congestion Model. Many studies document large and persistent earnings losses following job displacement events, of around 30% drop in earnings upon separation, with effects persisting even after twenty years (see, e.g., Davis and von Wachter, 2011). The leading explanations build on workers falling off the job ladder and the associated loss in job stability following a layoff (Jarosch, 2015; Jung and Kuhn, 2018). Importantly, these displacement costs are much larger in recessions than in booms, as documented in Davis and von Wachter (2011), a feature that is not yet well understood (see, e.g., Jung and Kuhn, 2018).

Countercyclical congestion can account for the countercyclicity of earnings losses from displacement. To highlight this result, we replicate the analysis in Davis and von Wachter (2011) in our model. Specifically, we compute the earnings trajectory of a cohort of separated workers, tak-

Figure 13: Path Dependence of Earnings Losses



Notes: Panel (a) plots the percentage point difference in earnings losses of displaced workers in recessions relative to booms in the data (Davis and von Wachter, 2011, Figure 4 Panel C), and in the congestion model. Panel (b) plots the effect of the business cycle (unemployment rate) at point of graduation on earnings over time in the data (Schwandt and von Wachter, 2019, Figure 2) and in the model. The model results are based on estimating the regression specification in Equation (29) using simulations from our baseline model.

ing into account their subsequent labor market transitions (out of and back into unemployment). We conduct this exercise under two scenarios: “booms” and “recessions.” Both are generated by separation shocks leading to an average 3.5 percentage point unemployment rate difference between troughs and peaks, the magnitude observed in the period 1980-2005 used in Davis and von Wachter (2011). We express the earnings of this cohort of “displaced workers” relative to a control group of “surviving” incumbents (i.e., those incumbent workers who did not get displaced at the time, but may fall into unemployment in the future). We also apply the model analogue of the sample restriction in Davis and von Wachter (2011), of at least three years of job tenure.

Figure 13 Panel (a) shows the *difference* in earnings losses from a job separation in recessions compared to booms for the model (blue solid line).³⁹ Workers displaced in a recession lose almost 15 percentage points more in earnings than workers displaced in booms. This difference fades only very gradually; even ten years after displacement, it remains at 5 percentage points. These model trajectories are close to the empirical ones estimated by Davis and von Wachter (2011), which we plot as the black dotted line.⁴⁰ Since our model assumes flexible wages (Nash bargaining with a high bargaining power of workers), wages are relatively sensitive to job-level productivity, which may explain the misalignment in the middle years of the response.

Costs of Graduating in a Recession. Business cycles also have strong effects on the life-time incomes of new graduates entering the labor market (see, e.g., Kahn, 2014; Oreopoulos, von

³⁹Since our model is calibrated such that all worker types have identical wages in steady state ($p_k = 1$ and hence $w_k = w$ for all k), it cannot speak to the *level* of displacement costs.

⁴⁰The empirical estimates of earnings losses from displacement in booms and recessions are presented in Figure 4 Panel (c), in Davis and von Wachter (2011). We plot the difference between the boom and recession estimates in our Figure 13 Panel (a).

Wachter, and Heisz, 2012; Schwandt and von Wachter, 2019). While our model does not contain a life-cycle dimension, we can proxy for it in our model by following newly hired workers entering the labor market with type $k = 1$. We estimate the following regression on model-simulated earnings paths of cohorts of newly hired workers, which mimics Equation (2) estimated on data in Schwandt and von Wachter (2019):

$$y_{g,t} = \alpha + \beta_e u_g + \lambda_g + \chi_t + \epsilon_{g,t}, \quad (29)$$

where $y_{g,t}$ is average earnings of a cohort in period t hired out of unemployment (“graduated”) in period g , u_g is the unemployment rate in period g (at the time of “graduation”), λ_g are graduation fixed effects, and χ_t are time fixed effects. The coefficients of interest are given by vector β_e , which captures the effect of the unemployment rate at the time of labor market entry on subsequent earnings, where $e = t - g$ captures time since graduation.

Figure 13 Panel (b) plots the β_e coefficients estimated on simulated data together with the empirical estimates from Schwandt and von Wachter (2019).⁴¹ The model closely matches the data, with a one percentage point increase in unemployment resulting in about a 3.5% drop in earnings on impact. These negative effects of entering the labor market during periods of heightened unemployment persist even ten years following labor market entry.

Mechanisms. Empirically, most of the proximate sources of these two types of countercyclical earnings losses are accounted for by declines in wage profiles (see, e.g., von Wachter, forthcoming), supporting the persistent cohort effects on productivity in our model. In addition, studies have found observed earnings losses to be associated with flows to lower wage firms (Schmieder, von Wachter, and Heining, 2019; Oreopoulos, von Wachter, and Heisz, 2012) and occupational switches or downgrading (Altonji, Kahn, and Speer, 2016; Huckfeldt, 2016). These patterns could be viewed as consistent with congestion manifesting itself as low-quality relative to high-quality firms absorbing the increase in UE hires.⁴²

6.3 Policy Insensitivity Despite Productivity-Driven Business Cycles

We finally revisit the dilemma formulated by Costain and Reiter (2008): a DMP model cannot simultaneously match the cyclicity of labor market variables in response to productivity shocks and the long-run sensitivity of these variables with respect to policies that affect job surplus, such as unemployment insurance (UI) benefits. In a cross-country analysis, Costain and Reiter (2008) estimate the semi-elasticity of the unemployment rate with respect to the replacement rate, $\epsilon_{u,b/\bar{w}} = \partial \ln u / \partial (b/\bar{w})$, to lie between 2 and 3.5. While the standard DMP model can replicate this semi-elasticity, it fails to generate sufficient volatility in labor market variables. By contrast,

⁴¹See Figure 2 in Schwandt and von Wachter (2019) for the empirical estimates in our Figure 13 Panel (b).

⁴²A complementary literature studies the destruction and creation of jobs by firm quality (Moscarini and Postel-Vinay, 2012; Haltiwanger, Hyatt, Kahn, and McEntarfer, 2018), and the countercyclicity of skill requirements (Modestino, Shoag, and Ballance, 2016).

the solution by Hagedorn and Manovskii (2008) to calibrate b to feature a small steady-state fundamental surplus (Ljungqvist and Sargent, 2017), generates sufficient volatility in labor market variables, but overstates their sensitivity to UI.⁴³

Returning to our model and starting from our baseline calibration, we increase the UI benefit level b by 1%, i.e., $b_{new} = 1.01b_{base}$, and recompute the steady state values for all the model variables. Following Costain and Reiter (2008), we calculate the semi-elasticity of unemployment with respect to the replacement rate as $\epsilon_{u,b/\bar{w}} = \frac{\ln u_{new} - \ln u_{base}}{(b_{new}/\bar{w}_{new}) - (b_{base}/\bar{w}_{base})} \approx 2.6$, a value well within the bounds reported by Costain and Reiter (2008). Hence, our framework *simultaneously* matches the high volatility of labor market variables and the lower sensitivity of these variables with respect to policy instruments. This is because our model generates labor market volatility through larger fluctuations in allocative productivity and surplus, while maintaining a relatively large steady-state match surplus, so it can afford small steady-state elasticities.

7 Conclusion

Recessions and their aftermath are times when *more* jobs are filled by recently unemployed workers. With limits on the economy’s capacity to absorb new hires brought about by diminishing returns in the types of jobs the unemployed fill—congestion in hiring—, the labor productivity of new hires falls by much more than average labor productivity, lowering further hiring incentives, and raising unemployment.

Our model with countercyclical congestion in hiring is consistent with a range of macroeconomic regularities. It performs well in explaining the volatility of labor market quantities while generating an empirically consistent strongly downward sloping Beveridge curve. The model does so while featuring a high fundamental surplus and not relying on wage rigidity. Our framework also rationalizes more cyclical wages of newly hired workers relative to the average wage, the countercyclical labor wedge, the countercyclical earnings losses upon job displacement and labor market entry, and a realistic long-run elasticity of the unemployment rate with respect to UI.

We close with questions our study leaves open. First, we have presented time series evidence consistent with congestion and reviewed corroborating cross sectional quasi-experimental studies—but we have not definitively quantified the degree of congestion in hiring. Moreover, while we discuss potential microfoundations for imperfect worker substitutability, we leave for future research to provide direct evidence on this mechanism and an identification of the parameter that guides it. Second, while our reinterpretation of existing evidence on the cyclical nature of wages has supported our productivity-based congestion mechanism, we have shown that congestion may emerge also from hiring costs or perhaps other factors. Third, we abstract from compositional effects (e.g., Mueller, 2017), which may raise the quality of the unemployed in recessions and hence offset the amplification of the productivity fluctuations of new hires we study. Fourth, our real

⁴³For empirical research on short-run effects of UI across US local labor markets, see Hagedorn, Karahan, Manovskii, and Mitman (2019); Chodorow-Reich, Coglianese, and Karabarbounis (2019); Boone, Dube, Goodman, and Kaplan (forthcoming).

model draws on two types of shocks (to productivity and the job separation rate), and we have sidestepped alternative drivers of business cycles; Broer, Druedahl, Harmenberg, and Öberg (2021) explore a model with demand constraints and aggregate demand shocks, and include a notion of congestion. Relatedly, these aggregate demand constraints may also generate congestion. Finally, our study suggests that factors and policies attenuating shifts in separations, such as firing taxes or furlough schemes, may attenuate shifts in the job finding rate.

References

- ABRAHAM, K. AND L. KATZ (1986): "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?" *Journal of Political Economy*, 94, 507–522.
- ALTONJI, J. G., L. B. KAHN, AND J. D. SPEER (2016): "Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success," *Journal of Labor Economics*, 34, S361–S401.
- BAKER, G., M. GIBBS, AND B. HOLMSTROM (1994): "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 109, 881–919.
- BARNICHON, R. (2010): "Building a Composite Help Wanted Index," *Economics Letters*, 109, 175–178.
- BEAUDRY, P. AND J. DINARDO (1991): "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99, 665–88.
- BILS, M. (1985): "Real wages over the business cycle: evidence from panel data," *Journal of Political Economy*, 93, 666–689.
- BILS, M., B. KAYMAK, AND K.-J. WU (2020): "Labor Substitutability among Schooling Groups," Mimeo.
- BILS, M., P. J. KLENOW, AND B. A. MALIN (2018): "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, 108, 1118–46.
- BLANCHARD, O. AND P. DIAMOND (1990): "The Cyclical Behavior of the Gross Flows of US Workers," *Brookings Papers on Economic Activity*, 1990, 85–155.
- BOONE, C., A. DUBE, L. GOODMAN, AND E. KAPLAN (forthcoming): "Unemployment insurance generosity and aggregate employment," *American Economic Journal: Policy*.
- BROER, T., J. DRUEDAHL, K. HARMENBERG, AND E. ÖBERG (2021): "The Unemployment-Risk Channel in Business-Cycle Fluctuations," *Working paper*.
- BURDA, M. AND C. WYPLOSZ (1994): "Gross Worker and Job Flows in Europe," *European Economic Review*, 38, 1287–1315.
- CAHUC, P., S. CARCILLO, AND T. LE BARBANCHON (2018): "The Effectiveness of Hiring Credits," *The Review of Economic Studies*, 86, 593–626.
- CAIRÓ, I., S. FUJITA, AND C. MORALES-JIMENEZ (2020): "The Cyclicity of Labor Force Participation Flows: The Role of Labor Supply Elasticities and Wage Rigidity," *Working Paper*.
- CARRILLO-TUDELLA, C., H. GARTNER, AND L. KAAS (2021): "Recruitment Policies, Job-Filling Rates and Matching Efficiency," Mimeo.
- CHARI, V., P. KEHOE, AND E. MCGRATTAN (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.
- CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2012): "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macro Annual 2012*.
- CHODOROW-REICH, G., J. COGLIANESE, AND L. KARABARBOUNIS (2019): "The Macro Effects of Unemployment Benefit Extensions: a Measurement Error Approach," *The Quarterly Journal of Economics*, 134, 227–279.
- CHODOROW-REICH, G. AND L. KARABARBOUNIS (2016): "The Cyclicity of the Opportunity Cost of Employment," *Journal of Political Economy*, 124, 1563–1618.
- CHODOROW-REICH, G. AND J. WIELAND (2020): "Secular Labor Reallocation and Business Cycles," *Journal of Political Economy*, 128, 2245–2287.
- CHRISTIANO, L., M. EICHENBAUM, AND M. TRABANDT (2016): "Unemployment and Business Cycles," *Econometrica*, 84, 1523–1569.
- (2020): "Why is Unemployment so Countercyclical?" *Working Paper*.
- COLES, M. AND A. MOGHADDASI KELISHOMI (2018): "Do Job Destruction Shocks Matter in the Theory of

- Unemployment?" *American Economic Journal: Macroeconomics*, 10, 118–36.
- COSTAIN, J. AND M. REITER (2008): "Business Cycles, Unemployment Insurance, and the Calibration of Matching Models," *Journal of Economic Dynamics and Control*, 32.
- DAVIS, S. AND T. VON WACHTER (2011): "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*, 1–72, autumn.
- DEN HAAN, W., C. HAEFKE, AND G. RAMEY (2005): "Turbulence and Unemployment in a Job Matching Model," *Journal of the European Economic Association*, 3, 1360–1385.
- DOERINGER, P. B. AND M. J. PIORE (1985): *Internal Labor Markets and Manpower Analysis*, ME Sharpe.
- DORAN, K., A. GELBER, AND A. ISEN (2020): "The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries," *Working Paper*.
- DUPRAZ, S., E. NAKAMURA, AND J. STEINSSON (2019): "A Plucking Model of Business Cycles," *Working Paper*.
- ELSBY, M., B. HOBIJN, AND A. ŞAHİN (2015a): "On the Importance of the Participation Margin for Labor Market Fluctuations," *Journal of Monetary Economics*, 72, 64–82.
- ELSBY, M., R. MICHAELS, AND D. RATNER (2015b): "The Beveridge Curve: A Survey," *Journal of Economic Literature*, 53, 571–630.
- ELSBY, M. W. L., B. HOBIJN, AND A. ŞAHİN (2013): "Unemployment Dynamics in the OECD," *The Review of Economics and Statistics*, 95, 530–548.
- ENGBOM, N. (2020): "Recruiting Talent During a Recession," *Working Paper*.
- EYIGUNGOR, B. (2010): "Specific Capital and Vintage Effects on the Dynamics of Unemployment and Vacancies," *American Economic Review*, 100, 1214–37.
- FACCINI, R. AND E. YASHIV (2020): "The Importance of Hiring Frictions in Business Cycles," IZA Discussion Paper, N. 12889.
- FERNALD, J. (2014): "A quarterly, utilization-adjusted series on total factor productivity," *Working Paper*.
- FERRARO, D. (2018): "The Asymmetric Cyclical Behavior of the US Labor Market," *Review of Economic Dynamics*, 30, 145–162.
- FUJITA, S. AND G. MOSCARINI (2017): "Recall and Unemployment," *American Economic Review*, 107, 3875–3916.
- FUJITA, S. AND G. RAMEY (2006): "The Cyclical Behavior of Job Loss and Hiring," *Working Paper*.
- (2007): "Job Matching and Propagation," *Journal of Economic Dynamics and Control*, 31, 3671–3698.
- (2009): "The Cyclical Behavior of Separation and Job Finding Rates," *International Economic Review*, 50, 415–430.
- GALÍ, J. AND T. VAN RENS (forthcoming): "The Vanishing Procyclicality of Labor Productivity," *The Economic Journal*.
- GATHMANN, C., I. HELM, AND U. SCHÖNBERG (2018): "Spillover Effects of Mass Layoffs," *Journal of the European Economic Association*, 18, 427–468.
- GERTLER, M., C. HUCKFELDT, AND A. TRIGARI (2020): "Unemployment Fluctuations, Match Quality and the Wage Cyclical Behavior of New Hires," *The Review of Economic Studies*.
- GERTLER, M. AND P. KARADI (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7.
- GERTLER, M. AND A. TRIGARI (2009): "Unemployment Fluctuations with Staggered Nash Wage Bargaining," *Journal of Political Economy*, 117, 38–86.
- GILCHRIST, S. AND E. ZAKRAJŠEK (2012): "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 102.
- GIUPPONI, G. AND C. LANDAIS (2020): "Subsidizing Labor Hoarding in Recessions: The Employment and Welfare Effects of Short Time Work," *Working Paper*.

- GOLOSOV, M. AND G. MENZIO (2020): "Agency Business Cycles," *Theoretical Economics*, 15, 123–158.
- GORODNICHENKO, Y. AND M. WEBER (2016): "Are Sticky Prices Costly? Evidence from the Stock Market," *American Economic Review*, 106, 165–99.
- GRIGSBY, J., E. HURST, AND A. YILDIRMAZ (forthcoming): "Aggregate Nominal Wage Adjustments: New Evidence from Administrative Payroll Data," *Working Paper*.
- GÓMEZ, V., A. MARAVALL, AND D. PEÑA (1999): "Missing Observations in ARIMA Models: Skipping Approach versus Additive Outlier Approach," *Journal of Econometrics*, 88, 341 – 363.
- GÜRKAYNAK, R. S., B. SACK, AND E. SWANSON (2005): "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review*, 95, 425–436.
- HAEFKE, C., M. SONNTAG, AND T. VAN RENS (2013): "Wage Rigidity and Job Creation," *Journal of Monetary Economics*, 60, 887 – 899.
- HAGEDORN, M., F. KARAHAN, I. MANOVSKII, AND K. MITMAN (2019): "Unemployment Benefits and Unemployment in the Great Recession: the Role of Macro Effects," *NBER Working Paper*.
- HAGEDORN, M. AND I. MANOVSKII (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49.
- (2013): "Job Selection and Wages over the Business Cycle," *American Economic Review*, 103, 771–803.
- HAGEDORN, M., I. MANOVSKII, AND S. STETSENKO (2016): "Taxation and Unemployment in Models with Heterogeneous Workers," *Review of Economic Dynamics*, 19, 161–189.
- HALL, R. (2005a): "The Amplification of Unemployment Fluctuations through Self-Selection," *NBER Working Paper*.
- (2005b): "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95, 50–65.
- (2017): "High Discounts and High Unemployment," *American Economic Review*, 107, 305–30.
- HALL, R. AND M. KUDLYAK (2020a): "The Inexorable Recoveries of US Unemployment," *NBER Working Paper*.
- (2020b): "Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?" *NBER Working Paper*.
- HALL, R. AND P. MILGROM (2008): "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98, 1653–74.
- HALTIWANGER, J. C., H. R. HYATT, L. B. KAHN, AND E. MCENTARFER (2018): "Cyclical Job Ladders by Firm Size and Firm Wage," *American Economic Journal: Macroeconomics*, 10, 52–85.
- HAZELL, J. AND B. TASKA (2020): "Downward Rigidity in the Wage for New Hires," *Working Paper*.
- HUCKFELDT, C. (2016): "Understanding the Scarring Effect of Recessions," *Working Paper*.
- JACOBSON, L., R. LALONDE, AND D. SULLIVAN (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 685–709.
- JÄGER, S., B. SCHOEFER, S. YOUNG, AND J. ZWEIMÜLLER (2020): "Wages and the Value of Nonemployment," *The Quarterly Journal of Economics*, 135, 1905–1963.
- JAROSCH, G. (2015): "Search for Job Security and the Consequences of Job Loss," *Working Paper*.
- JEONG, H., Y. KIM, AND I. MANOVSKII (2015): "Price of Experience," *American Economic Review*, 105, 784–815.
- JUNG, P. AND M. KUHN (2018): "Earnings Losses and Job Mobility over the Lifecycle," *Journal of the European Economic Association*, 17, 678 – 724.
- JURADO, K., S. LUDVIGSON, AND S. NG (2015): "Measuring Uncertainty," *American Economic Review*, 105.
- KAHN, L. (2014): "The Longterm Labor Market Consequences of Graduating College in a Bad Economy," *Labour Economics*, 17, 206–223.
- KARABARBOUNIS, L. (2014): "The Labor Wedge: MRS vs MRN," *Review of Economic Dynamics*, 17, 303–316.

- KATZ, L. AND B. MEYER (1990): "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes," *The Quarterly Journal of Economics*, 105, 973–1002.
- KATZ, L. AND K. MURPHY (1992): "Changes in Relative Wages, 1963-87: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.
- KEHOE, P., P. LOPEZ, V. MIDRIGAN, AND E. PASTORINO (2019): "Asset Prices and Unemployment Fluctuations," *NBER Working Paper*.
- KROFT, K., F. LANGE, AND M. NOTOWIDIGDO (2013): "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 128.
- KRUSELL, P., T. MUKOYAMA, R. ROGERSON, AND A. ŞAHİN (2017): "Gross Worker Flows Over the Business Cycle," *American Economic Review*, 107, 3447–76.
- KUDLYAK, M. (2014): "The Cyclicalities of the User Cost of Labor," *Journal of Monetary Economics*, 68, 53–67.
- LANDAIS, C., P. MICHAILLAT, AND E. SAEZ (2018): "A Macroeconomic Approach to Optimal Unemployment Insurance: Theory," *American Economic Journal: Economic Policy*, 10, 152–81.
- LAZEAR, E. AND P. OYER (2004): "Internal and External Labor Markets: A Personnel Economics Approach," *Labour Economics*, 11.
- LILIE, D. (1982): "Sectoral Shifts and Cyclical Unemployment," *Journal of Political Economy*, 90, 777–793.
- LJUNGQVIST, L. AND T. SARGENT (1998): "The European Unemployment Dilemma," *Journal of Political Economy*, 106, 514–550.
- (2004): "European Unemployment and Turbulence Revisited in a Matching Model," *Journal of the European Economic Association*, 2, 456–468.
- (2017): "The Fundamental Surplus," *American Economic Review*, 107.
- MANGIN, S. AND B. JULIEN (2020): "Efficiency in Search and Matching Models: A Generalized Hosios Condition," *Working Paper*.
- MERCAN, Y. AND B. SCHOEFER (2020): "Jobs and Matches: Quits, Replacement Hiring, and Vacancy Chains," *American Economic Review: Insights*, 2, 101–24.
- MIAN, A. AND A. SUFI (2014): "What Explains the 2007–2009 Drop in Employment?" *Econometrica*, 82, 2197–2223.
- MICHAILLAT, P. (2012): "Do Matching Frictions Explain Unemployment? Not in Bad Times," *American Economic Review*, 102, 1721–50.
- MITMAN, K. AND S. RABINOVICH (2020): "Do Unemployment Benefit Extensions Explain the Emergence of Jobless Recoveries?" *Working Paper*.
- MODESTINO, A. S., D. SHOAG, AND J. BALLANCE (2016): "Downskilling: Changes in Employer Skill Requirements over the Business Cycle," *Labour Economics*, 41, 333–347.
- MORTENSEN, D. AND C. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *The Review of Economic Studies*, 61, 397–415.
- MOSCARINI, G. AND F. POSTEL-VINAY (2012): "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment," *American Economic Review*, 102, 2509–39.
- MUELLER, A. (2017): "Separations, Sorting, and Cyclical Unemployment," *American Economic Review*, 107, 2081–2107.
- OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4, 1–29.
- PISSARIDES, C. A. (2009): "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica*, 77, 1339–1369.
- ROMER, C. D. AND D. H. ROMER (2004): "A New Measure of Monetary Shocks: Derivation and Implications,"

- American Economic Review*, 94, 1055–1084.
- SCHMIEDER, J., T. VON WACHTER, AND J. HEINING (2019): “The Costs of Job Displacement over the Business Cycle and its Sources: Evidence from Germany,” *Working Paper*.
- SCHOEFER, B. (2021): “The Financial Channel of Wage Rigidity,” *NBER Working Paper*.
- SCHWANDT, H. AND T. VON WACHTER (2019): “Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets,” *Journal of Labor Economics*, 37, S161–S198.
- SHIMER, R. (2004): “The Consequences of Rigid Wages in Search Models,” *Journal of the European Economic Association*, 2, 469–479.
- (2005): “The Cyclical Behavior of Equilibrium Unemployment and Vacancies,” *American Economic Review*, 95, 25–49.
- (2009): “Convergence in Macroeconomics: The Labor Wedge,” *American Economic Journal: Macroeconomics*, 1, 280–97.
- (2010): *Labor Markets and Business Cycles*, Princeton University Press.
- (2012): “Reassessing the Ins and Outs of Unemployment,” *Review of Economic Dynamics*, 15, 127 – 148.
- SILVA, J. AND M. TOLEDO (2013): “The Unemployment Volatility Puzzle: The Role of Matching Costs Revisited,” *Economic Inquiry*, 51, 836–843.
- SILVA, J. I. AND M. TOLEDO (2009): “Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment,” *Macroeconomic Dynamics*, 13, 76–96.
- UHLIG, H. (2005): “What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure,” *Journal of Monetary Economics*, 52, 381–419.
- VON WACHTER, T. (forthcoming): “The Persistent Effects of Initial Labor Market Conditions for Young Adults and their Sources,” *Journal of Economic Perspectives*.
- WIELAND, J. F. AND M.-J. YANG (2020): “Financial Dampening,” *Journal of Money, Credit and Banking*, 52, 79–113.

Online Appendix of:
A Congestion Theory of Unemployment Fluctuations
Yusuf Mercan, Benjamin Schoefer and Petr Sedláček

Contents

A Empirical Appendix: Construction of Worker Flows and Transition Probabilities, and Robustness	52
B Further Details on Identification of Congestion	66
C A Generalization of the Baseline Model: Types vs. Inputs	71
D Solution Method	72
E Details of the Baseline Parameterization: Homogeneous Steady State Marginal Products Across Types	74
F Alternative Calibration: Small Surplus/"High b"	75
G Robustness of Model Performance to the Number of Worker Types	78
H Business Cycle Statistics Including Full Correlation Matrices	81
I Alternative Calibration: Accounting for Labor Force Participation and Matching EU Flows	84
J Deriving the Iso-congestion Curve	88
K Alternative Mechanism: Convex Hiring Costs	91
L Historical Decomposition: Additional Material	93

List of Tables

A1	Discrete vs. Time-Aggregation Adjusted Worker Transitions	56
A2	Business Cycle Properties: Alternative Smoothing Parameter	61
A3	Benchmarks: Elasticity of Substitution Parameters Estimated in the Literature	67
A4	Separation Shocks and Other Disturbances: Adjusted R-squared	70
A5	Business Cycle Properties: No-Congestion, Low-Surplus Model	77
A6	Business Cycle Properties in the Data	81
A7	Business Cycle Properties in the No-Congestion Model without Separation Shocks	81
A8	Business Cycle Properties in the No-Congestion Model with Separation Shocks	82
A9	Business Cycle Properties in the Congestion Model—Baseline (Matching UE/E)	82
A10	Business Cycle Properties in the Congestion Model—Robustness (Matching EU & Participation)	83
A11	Business Cycle Properties: Convex Hiring Cost Model	92
A12	Historical Decomposition of Unemployment: Model and Counterfactuals	95

List of Figures

A1	The Countercyclicality of Hiring out of Unemployment: Okun's Laws and Elasticities	52
A2	Comparing Discrete and Time Aggregation Adjusted UE Flows	57
A3	Comparing Discrete and Time Aggregation Adjusted Flow Probabilities	58
A4	Comparing Discrete and Time Aggregation Adjusted UE Flows (Shimer 2012)	59
A5	Countercyclicality of the Employment Share with Nonemployment Past Year	60
A6	The Countercyclicality of New Hire Share: CPS Worker Flows	61
A7	Cyclicality of Share of New Hires in Employment: CPS Worker Flows	62
A8	Cyclicality of New Hires: CPS Worker Flows	63
A9	The Countercyclicality of Unemployment-to-Employment Flows	64
A10	Cyclicality of UE Flows in the OECD	65
A11	IRF of Unemployment Rate to a Separation Rate Shock: Data and Models	66
A12	IRF of Labor Productivity to a Separation Rate Shock: Data and Models	66
A13	Separation Shocks: Baseline and Adjusted for Identified Shocks	70
A14	Relative Worker Productivities in the Congestion Model	74
A15	Impulse Responses to a Separation Shock: No-Congestion, Low-Surplus Model	76
A16	Volatility of unemployment, slope of the Beveridge curve and model fit vs K	78
A17	Separation Rate: Measured and Implied	85
A18	Impulse Responses to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows	87
A19	Iso-congestion Curves	90
A20	Impulse Responses to a Separation Shock: Convex Hiring Cost Model	92

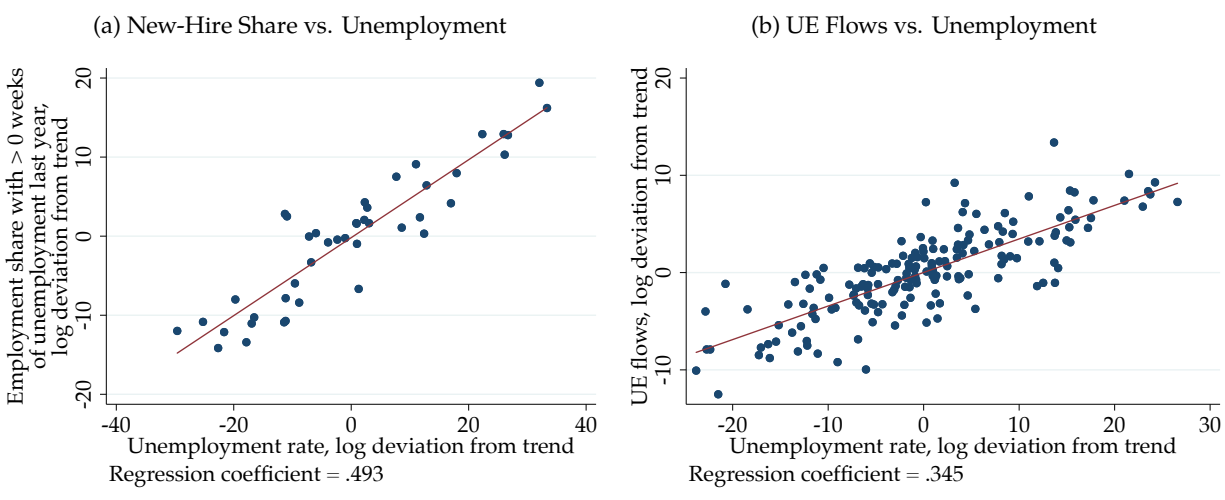
A21 Time Paths of Labor Market Variables	94
A22 Unemployment Components	95

A Empirical Appendix: Construction of Worker Flows and Transition Probabilities, and Robustness

In Appendix A.1, we elaborate on our construction of worker flows and measurement of transition probabilities. The rest of the sections provide more details for each of the robustness exercises.

We organize our robustness checks by comparing results to the baseline definition of hires, which we represent as Okun's laws (the elasticity of new hires' outcomes with respect to the unemployment rate, both in log deviations) in Figure A1.

Figure A1: The Countercyclicality of Hiring out of Unemployment: Okun's Laws and Elasticities



Notes: Panel (a) plots log deviations in the share of employed with some unemployment spell in the preceding year against log deviations in the unemployment rate. The time series are HP filtered with a smoothing parameter of 100. Source: CPS March Supplement (ASEC). Panel (b) plots log deviations in UE flows against log deviations in the unemployment rate. The time series are HP filtered with a smoothing parameter of 1,600. Source: CPS monthly files. We also report elasticities (the linear regression coefficients) for both panels.

Below is a summary list of the robustness checks we undertake:

Time Aggregation Adjustment. For consistency with the discrete time model presented in the main text, the empirical transition rates are not adjusted for time-aggregation bias. In other words, initially employed workers may separate into unemployment and transition back into employment within the period—as in the CPS ASEC definition of asking the end-of-period employed about potential unemployment spells during the period. In Appendix A.2, we find very similar results for the cyclical behavior of these UE flows adjusted for such time aggregation.

Unemployment vs. Nonemployment. In Appendix A.3, in Figures A5 to A8, we replicate Figure 1 by considering the nonemployment (comprising the unemployed and out of the labor force) rather than the unemployment history of the employed, and find qualitatively similar cyclical patterns. While the countercyclicality of NE-hire share in employment exhibits a weaker Okun's

law, our model results would remain unaffected, since the model parameterization would simply require us to estimate a stronger degree of congestion in order to match our empirical calibration targets. In a model extension in Appendix I, we consider flows in and out of the labor force.

Alternative Detrending. In our main specification, we use the conventional smoothing parameter for quarterly data of 1,600 when studying worker flows and transition probabilities (see, e.g., Fujita and Ramey, 2009). Shimer (2005, 2012) instead chooses a smoothing parameter of 10^5 and accordingly attributes more of the time series variation to cyclical fluctuations. In Appendix A.4, we show that our results are robust to this alternative smoothing parameter. Most importantly, the elasticity of UE flows with respect to the unemployment rate stays unchanged (0.348 vs 0.345) as does the employment share of new hires out of unemployment (0.433 vs 0.432).

OECD Evidence. Appendix A.5 shows that countercyclical UE flows are a feature across the OECD; this fact has been documented (but not studied as a source of amplification) by, e.g., Blanchard and Diamond (1990); Burda and Wyplosz (1994); Fujita and Ramey (2009); Elsby, Hobijn, and Şahin (2013).

A.1 Baseline: Discrete Data

We use the Current Population Survey (CPS) to measure worker flows. The CPS has a rotating-panel design, in which households are surveyed for four consecutive months, then they rotate out for eight months and then are surveyed for another four months, after which they permanently leave the sample. This structure allows us to match at most three-fourths of the sample in one month to the next. In practice, the matching rate is below 75% due to the temporary absence of individuals from their residence or a household moving out of their address. This phenomenon is referred to as *margin error*.

We start with the monthly micro data covering January 1976 to December 2019. We restrict our sample to civilians age 15 and above. We categorize each individual in each month t into one of three employment states: employed (E), unemployed (U) and out of the labor force (O). We use final person-level weights to calculate the stock of employed, unemployed and non-participants, $E(t)$, $U(t)$, $O(t)$, for each month t .

Using individual identifiers (using the CPS samples provided by IPUMS and its unique individual ID, CPSIDP—which uses rotation groups, household identifiers, individual line numbers, race, sex, and age to identify individuals—we calculate individual-level transition events between consecutive months. We again use the current month person-level weights to calculate the total count of worker flows. Let $Z_{ij}(t)$ denote worker flows: the mass of workers in employment state i in month $t - 1$ that are observed in employment state j in month t for $i, j \in \{E, U, O\}$.

To correct for margin error, we make the common *missing at random* (MAR) assumption, which omits missing observations and reweights the measured flows. We adjust our time series by

reweighting the measured flows $Z_{ij}(t)$ for $i, j \in \{E, U, O\}$ as follows:

$$\mu_{ij}(t) = \frac{E(t) + U(t) + O(t)}{\sum_i \sum_j Z_{ij}(t)} Z_{ij}(t).$$

The numerator is the worker population implied by measured stocks and the denominator is the population implied by total measured flows, including workers whose employment states do not change. In practice, we construct $\mu_{ij}(t)$ for males and females separately, and then sum them to arrive at our aggregate measure of worker flows adjusted for margin error.

For a number of months in the CPS, it is impossible to match individuals over time. The raw flow series also exhibit several extreme jumps. To deal with missing values and outliers, we follow the approach outlined in Fujita and Ramey (2006) and use the procedure called Time Series Regression with ARIMA Noise, Missing Observations and Outliers (TRAMO, Gómez, Maravall, and Peña, 1999). We let TRAMO detect additive and transitory outliers using a pre-determined t-test critical level set to 4. Finally, we seasonally adjust the time series using the X-ARIMA-12 procedure developed by the US Census Bureau.

Finally, we calculate the discrete-time job finding and separation probabilities as

$$\begin{aligned} f_t &= \frac{\mu_{UE}(t)}{U(t-1)} \\ \delta_t &= \frac{\mu_{EU}(t)}{E(t-1)}, \end{aligned} \tag{A1}$$

which simply capture the share of unemployed (employed) workers in month $t-1$ who are observed to be employed (unemployed) in month t .

To sum up, the figures we present and our calibration targets in the model are based on our margin-error adjusted flow time series (under the MAR assumption) $\mu_{ij}(t)$, whose missing values and outliers are corrected by the TRAMO procedure, and are seasonally adjusted using the X-ARIMA-12 procedure.

A.2 Robustness: Time-Aggregation-Adjusted Data

Our preferred measure of worker flows is based on discrete time and hence is subject to a specific form of time aggregation bias: drawing on the CPS panel structure, we obtain worker flows by following initially unemployed workers that move into employment by the end of the period (are employed the beginning of next period). One type of transition we miss in this discrete-time approach is that initially employed workers may separate within the period and find a job again, akin to the issues laid out in Shimer (2005).

In this section, we compare the properties of UE flows based on our measurement approach in the main text to a one accounting for time-aggregation bias. Our object of interest is the total number of UE flows within the period, into jobs that remain active until the end of the period, mirroring our definition using the CPS ASEC in Section 2. We also confirm that our time series

replicate those reported by Shimer (2012).

Our Method. We draw on Fujita and Ramey (2006), who provide expressions for time-aggregation-adjusted gross worker flows, whereas our interest is in within-period cumulative UE flows that remain active through the end of the period.

First, we start with the monthly job finding f_t and separation δ_t probabilities, whose measurement are described in Appendix A.1, underlying the analysis in the main text.

Second, we compute the monthly job finding and separation hazards, \widehat{f}_t and $\widehat{\delta}_t$, solving the following system of equations:

$$\begin{aligned}\delta_t &= u_{ss,t}(1 - e^{-\widehat{f}_t - \widehat{\delta}_t}) \\ f_t &= (1 - u_{ss,t})(1 - e^{-\widehat{f}_t - \widehat{\delta}_t}),\end{aligned}\tag{A2}$$

where $u_{ss,t} = \widehat{\delta}_t / (\widehat{\delta}_t + \widehat{f}_t)$ is the steady-state approximation to the unemployment rate implied by the contemporaneous hazard rates. The unemployment law of motion in continuous time is

$$U_{t-1+\tau} = \frac{(1 - e^{-(\widehat{f}_t + \widehat{\delta}_t)\tau})\widehat{\delta}_t}{\widehat{f}_t + \widehat{\delta}_t} L_{t-1} + e^{-(\widehat{f}_t + \widehat{\delta}_t)\tau} U_{t-1},\tag{A3}$$

for $\tau \in [0, 1)$ and where L_t is the size of the labor force in month t .

Third, we calculate the number of employed workers at the end of month t who had any unemployment spell during t —which we then compare to the discrete-time-based UE flows. As an intermediate step, we consider the probability of not losing a job, from $t - 1 + \tau$ until t for $\tau \in [0, 1)$, conditional on having a job at t .¹ This probability is given by

$$\lim_{\Delta \rightarrow 0} \left(1 - \Delta \widehat{\delta}_t\right)^{\frac{1-\tau}{\Delta}} = e^{-\widehat{\delta}_t(1-\tau)}.\tag{A4}$$

Using this intermediate result, UE flows during month t , adjusted for time aggregation in that they also count within-period EUE transitions, are given by

$$UE_t = \int_0^1 \underbrace{\widehat{f}_t}_{\text{Find job}} \underbrace{U_{t-1+\tau}}_{\text{Number of unemployed}} \underbrace{e^{-\widehat{\delta}_t(1-\tau)}}_{\text{Do not lose job}} d\tau.\tag{A5}$$

Finally, using Equation (A3), we can integrate out the above expression to obtain UE flows adjusted

¹Therefore, our results do not study cycles such as “E(UEUEU)E” transitions during the period. These are comparatively tiny compared to the first-order flows stemming from the initially employed losing their job during the period, becoming reemployed, and not losing that first-found job again.

Table A1: Discrete vs. Time-Aggregation Adjusted Worker Transitions

UE flows	Discrete	Time-aggregation adjusted
Standard deviation	0.045	0.040
Autocorrelation	0.671	0.574
Correlation Matrix		
Discrete	1	
Time-aggregation adjusted	0.983	1

Notes: The table compares the time series properties of UE flows based on our discrete time measurement approach used in the main text to a version corrected for time-aggregation bias. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1,600.

for time aggregation bias:

$$UE_t = \widehat{f}_t L_{t-1} e^{-\widehat{\delta}_t} \left(u_{ss,t} \frac{e^{\widehat{\delta}_t} - 1}{\widehat{\delta}_t} + \left(\frac{U_{t-1}}{L_{t-1}} - u_{ss,t} \right) \frac{1 - e^{-\widehat{f}_t}}{\widehat{f}_t} \right). \quad (\text{A6})$$

Table A1 summarizes the properties of the discrete-time and time-aggregation-adjusted series. The two time series have extremely similar standard deviations and autocorrelations, and are nearly perfectly correlated.

Figure A2 Panel (a) reports the time series of UE flows in our baseline definition based on discrete time measurement, along with the time-aggregation-adjusted time series. Panel (b) shows the Okun's law, such that the elasticity of UE flows adjusted for time aggregation bias with respect to the unemployment rate is 0.265, similar to the elasticity arising from the discrete-time approach in Figure A1 Panel (b), where we estimated an only slightly higher elasticity of 0.345. Hence, our congestion dynamics are robust to time-aggregation adjustment, i.e., to counting within-period EUE flows in addition to the transitions into employment for the initially unemployed.

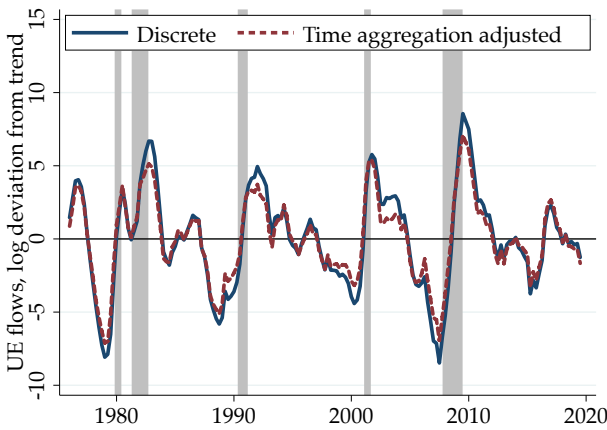
To gauge the accuracy of the time-aggregation adjusted hazard rates, \widehat{f} and $\widehat{\delta}$, in Panel (c) of Figure A2, we further plot the actual unemployment rate as well as its steady-state approximation $u_{ss,t}$. The steady-state approximation tracks the actual time series closely, lending credibility on the measurement exercise in this section.

Comparison to Shimer (2012). To further highlight the robustness and validity of our empirical analysis, we compare our preferred worker transition probabilities to the ones reported in Shimer (2012). Panel (a) in Figure A3 plots the employment-to-unemployment probability used in the main text and compares that to the same monthly probability adjusted for time aggregation bias provided by Shimer (2012). Panel (b) does the same for unemployment-to-employment flows. While the time-aggregation adjusted probabilities are higher in levels, their cyclical behavior closely tracks the underlying discrete-time probabilities that we use in our main analysis (Panels (c) and (d)).

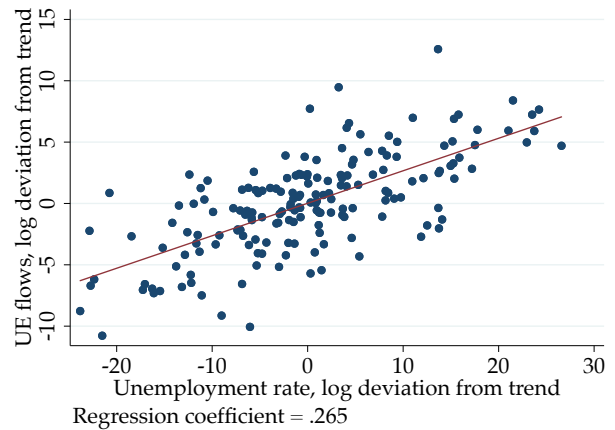
While Shimer (2012) does not report properties of UE flows in the paper, the similarity of the cyclical behavior of the transition rates also implies that the UE flows implied by the Shimer (2012)

Figure A2: Comparing Discrete and Time Aggregation Adjusted UE Flows

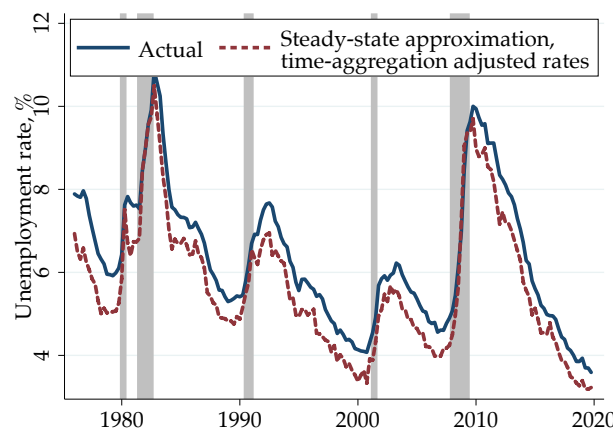
(a) UE Flows: Discrete vs. Time Aggregation Adjusted



(b) UE Flows vs. Unemployment Rate



(c) Unemployment Rate

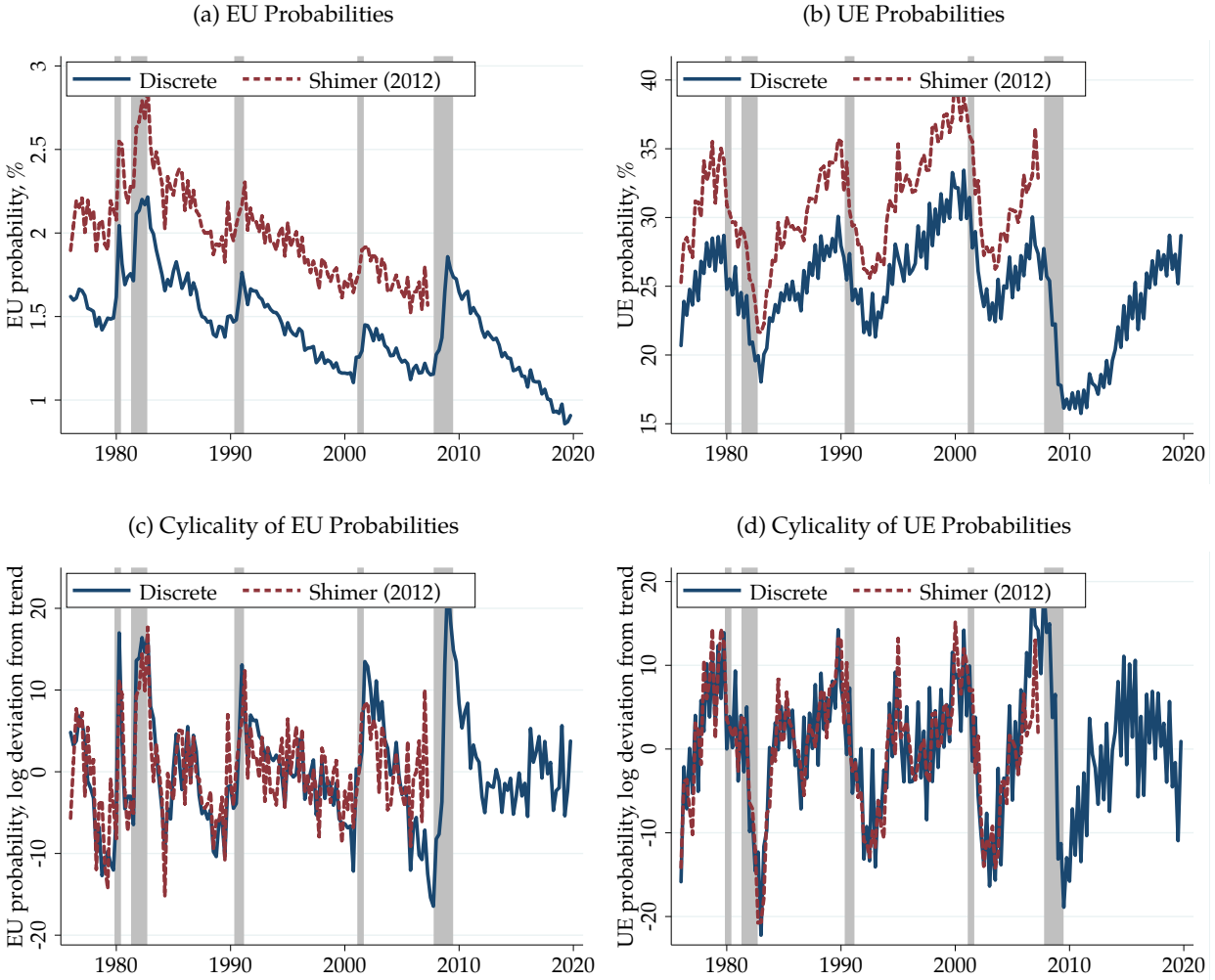


Notes: The figure shows robustness of the UE flows to time aggregation bias adjustment. Panel (a) reports the time series of UE flows in our baseline definition based on discrete time, along with the time-aggregation-debiased time series. Panel (b) is a scatter plot of UE flows adjusted for time aggregation bias against the unemployment rate. Panel (c) plots the actual unemployment rate and its steady-state approximation based on time-aggregation adjusted hazard rates, \hat{f} and $\hat{\delta}$. All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1,600. Source: CPS monthly files.

data would be similarly countercyclical.² For comparison with our main analysis, we calculate the

²Shimer (2012) does not present UE flows, but focuses on transition rates. In the discussion of the prior evidence, he writes: “In fact, even after accounting for time aggregation, the decline in the job finding probability almost exactly offsets the increase in the number of unemployed workers at business cycle frequencies, so the number of unemployed workers who find a job in a month shows little cyclicity” (page 145). Our reading is that this statement likely assesses the amplitude of log UE flows (i.e., percent deviations from trend) when compared with the amplitude of percent deviations from trend of the transition rates and probabilities, rather than a different conclusion of the qualitative nature about the countercyclicality of UE flows.

Figure A3: Comparing Discrete and Time Aggregation Adjusted Flow Probabilities



Notes: Panel (a) compares the EU probability used in the main text to its time-aggregation adjusted counterpart provided by Shimer (2012) allowing for flows between employment, unemployment and inactivity. Panel (b) does the same for UE probability. Panels (c) and (d) plot the log deviations of these probabilities from their respective trends. The series are logged and HP-filtered using a smoothing parameter of 1,600. Source: CPS monthly files.

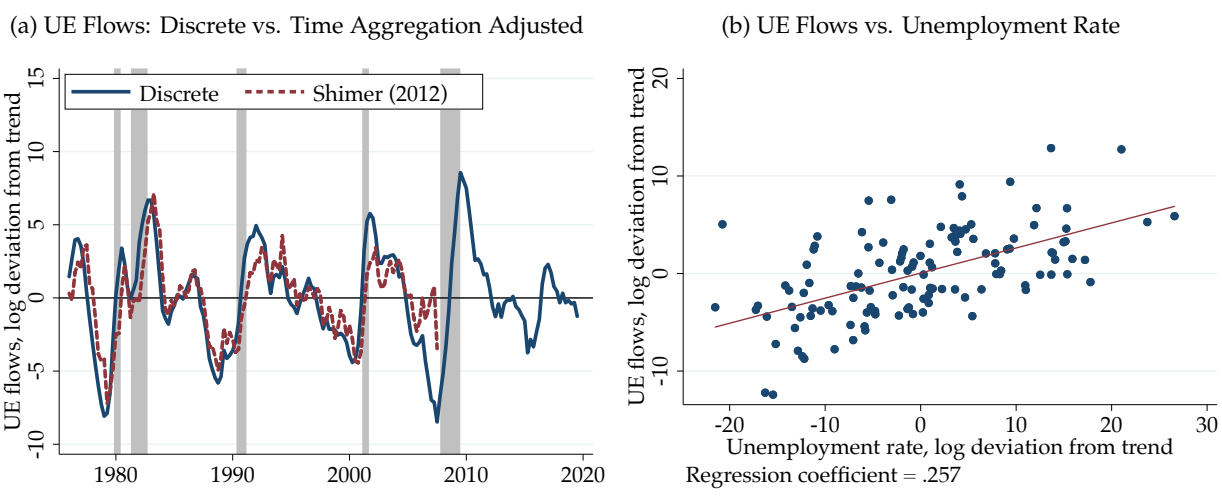
time-aggregation-adjusted UE flows simply as

$$UE_t^{\text{Shimer}} = U_{t-1} \Lambda_{ue,t}, \quad (\text{A7})$$

where $\Lambda_{ue,t}$ is the monthly probability of a UE flow provided by Shimer (2012).

Figure A4 compares our baseline measure of UE flows to the one based on Shimer (2012). Panel (a) plots the cyclical component of UE flows over time and shows that UE flows adjusted for time aggregation in a three-state model also exhibit strong countercyclicality. Panel (b) quantifies this countercyclicality: the elasticity of UE flows with respect to the unemployment rate is 0.257, slightly lower than the elasticity we report in the main text.

Figure A4: Comparing Discrete and Time Aggregation Adjusted UE Flows (Shimer 2012)



Notes: The figure shows robustness of the UE flows to time aggregation bias adjustment allowing for worker flows between three labor market states. Panel (a) reports the time series of UE flows in our baseline definition based on discrete time, along with the time-aggregation-debiased time series. Panel (b) is a scatter plot of UE flows adjusted for time aggregation bias against the unemployment rate. All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1,600. Source: CPS monthly files.

A.3 Robustness to Nonemployment rather than Unemployment, i.e., Including Previous Non-participants in New Hires

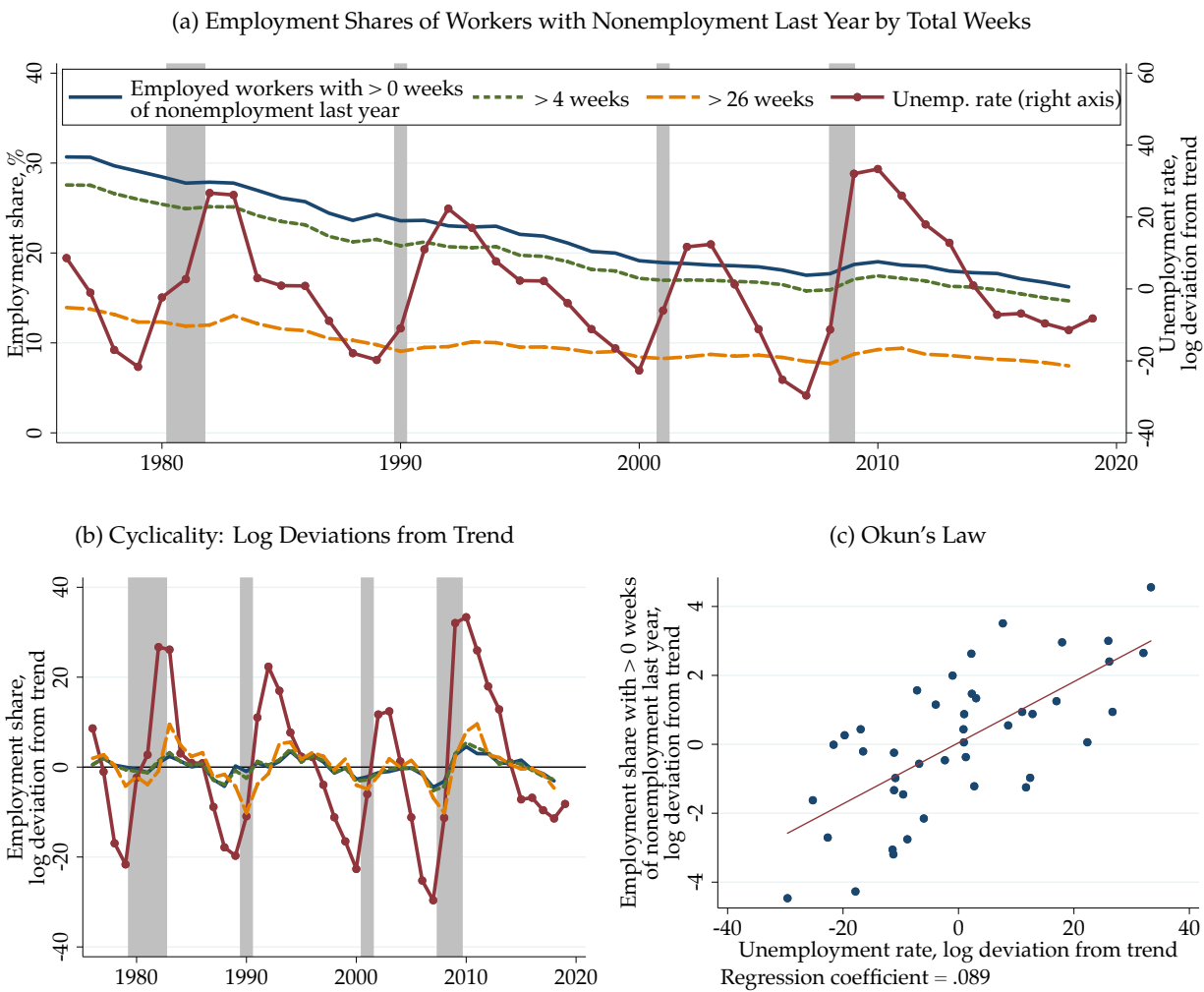
Below in Figures A5 to A8, we provide similar figures to those presented in the main text by including workers who flow into employment from non-participation. In Figure A5, we replicate Figure 1 by considering the nonemployment (comprising unemployment and out of the labor force) rather than the unemployment history of the employed, and find qualitatively similar cyclical patterns. While the countercyclicality of NE-hire share in employment exhibits a weaker Okun's law, our model results would remain unaffected, since the model parameterization would simply require us to estimate a stronger degree of congestion (lower σ) in order to match our empirical calibration targets.

A.4 Alternative HP Smoothing Parameter

In the main text, we report business cycle statistics based on HP-filtered time series with a smoothing parameter of 1,600, typically used for quarterly data. In this section, we instead use a smoothing parameter of 10^5 —preferred by Shimer (2005, 2012)—to report business-cycle statistics.

Table A2 reports the standard deviations, auto- and cross-correlations of the HP-filtered time series we present in the main text. With a smoothing parameter more aggressively penalizing movements in the trend components in the time series, the standard deviations of the variables around these trends become considerably higher. The cross-correlations between f , δ and UE/E

Figure A5: Countercyclicity of the Employment Share with Nonemployment Past Year



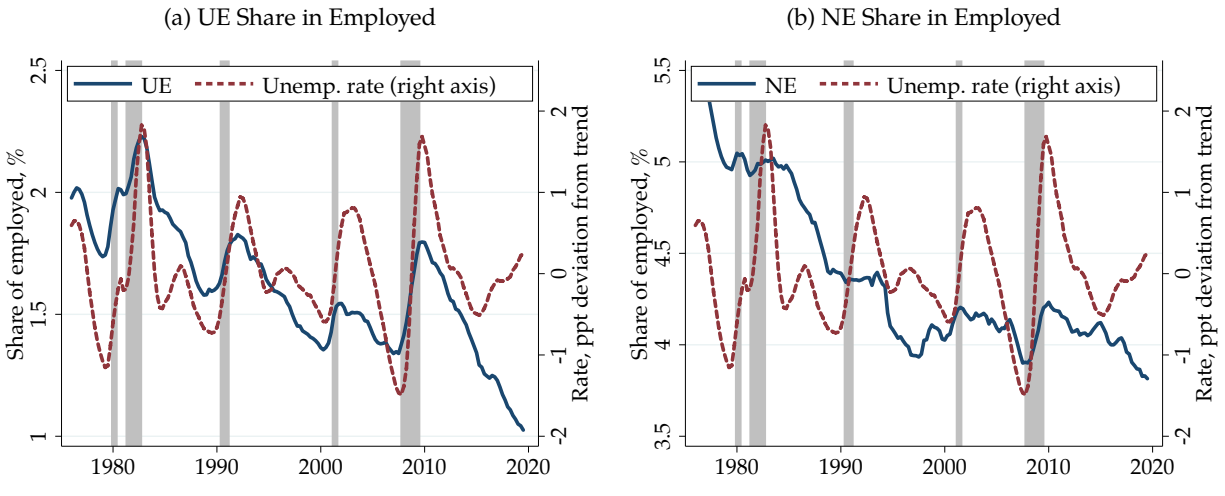
Notes: The figure replicates Figure 1, but instead conditions on *nonemployment* duration, i.e., we also include labor market states where a worker might be out of the labor force. Panel (a) plots the share of employed workers who have undergone a nonemployment spell in the preceding calendar year for different nonemployment durations. Panel (b) plots their log deviations from trend. Panel (c) reports the scatter plot of the detrended time series. The time series are HP filtered with a smoothing parameter of 100. Shaded regions denote NBER-dated recessions. Source: CPS March Supplement (ASEC).

become if anything even more pronounced.³

Most importantly, Figure A9 presents scatter plots of UE flows and shares against the unemployment rate, respectively, under this alternative smoothing parameter. The elasticity of UE flows with respect to the unemployment rate is almost identical to the one we present in Figure A1 Panel (b) (0.348 vs 0.345). Likewise, the elasticity of new-hire share in employment to the unemployment

³The correlation of average labor productivity with the job finding rate (unemployment rate) turns slightly negative (positive), likely due to the inclusion of additional years compared to Shimer (2005), and consistent with our aforementioned comment that ALP is not an obvious cyclical driver (see, e.g., Shimer, 2005; Mitman and Rabinovich, 2020; Galí and Van Rens, forthcoming).

Figure A6: The Countercyclicality of New Hire Share: CPS Worker Flows



Notes: Panel (a) plots the share of UE hires in employment. Panel (b) plots NE flows in the share of employed. All time series are based on quarterly averages of monthly data and for visual clarity are smoothed by taking centered four-quarter moving averages. Both panels also plot the percentage point deviation of unemployment rate from its trend on a secondary axis. Shaded regions denote NBER-dated recessions. Source: CPS monthly files.

Table A2: Business Cycle Properties: Alternative Smoothing Parameter

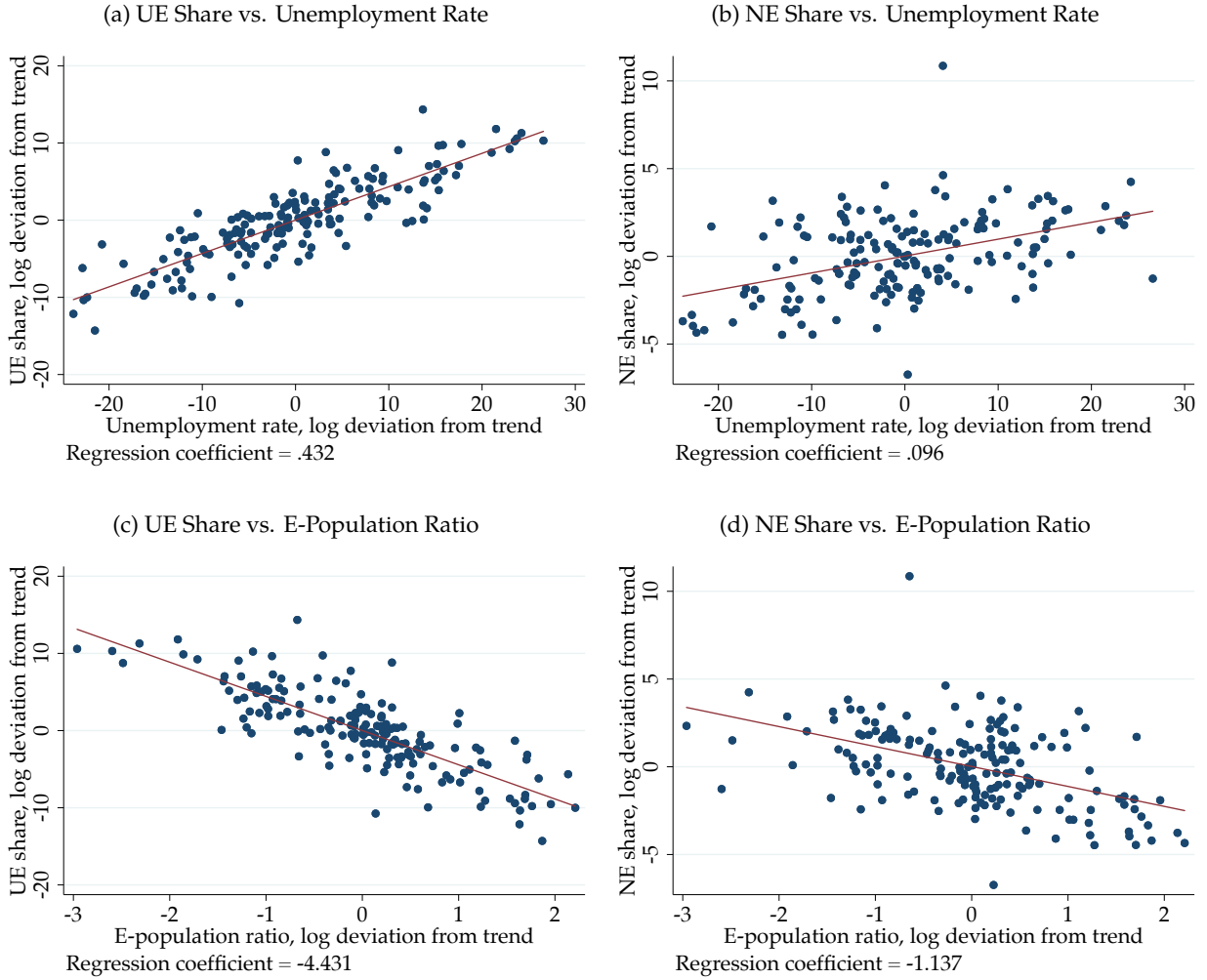
	ALP	f	δ	u	v	θ	UE/E
Standard deviation	0.017	0.093	0.108	0.190	0.198	0.376	0.116
Autocorrelation	0.897	0.950	0.904	0.970	0.957	0.962	0.933
Correlation matrix							
ALP	1						
f	-0.061	1					
δ	-0.179	-0.859	1				
u	0.015	-0.975	0.919	1			
v	0.050	0.831	-0.830	-0.851	1		
θ	0.038	0.906	-0.877	-0.928	0.978	1	
UE/E	0.113	-0.888	0.783	0.930	-0.718	-0.818	1

Notes: ALP , f , δ , u , θ and UE/E indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and share of new hires in employment. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with the alternative smoothing parameter of 10^5 rather than 1,600.

rate stays unchanged compared to the one reported in Figure A7 Panel (a) (0.433 vs 0.432).

We conclude that our key facts are robust to an alternative smoothing parameter of 10^5 preferred by Shimer (2005, 2012).

Figure A7: Cyclicity of Share of New Hires in Employment: CPS Worker Flows



Notes: The figure plots different measures of new-hire share in employment (UE or NE) against employment measures (unemployment rate or employment-population ratio). All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1,600. Source: CPS monthly files.

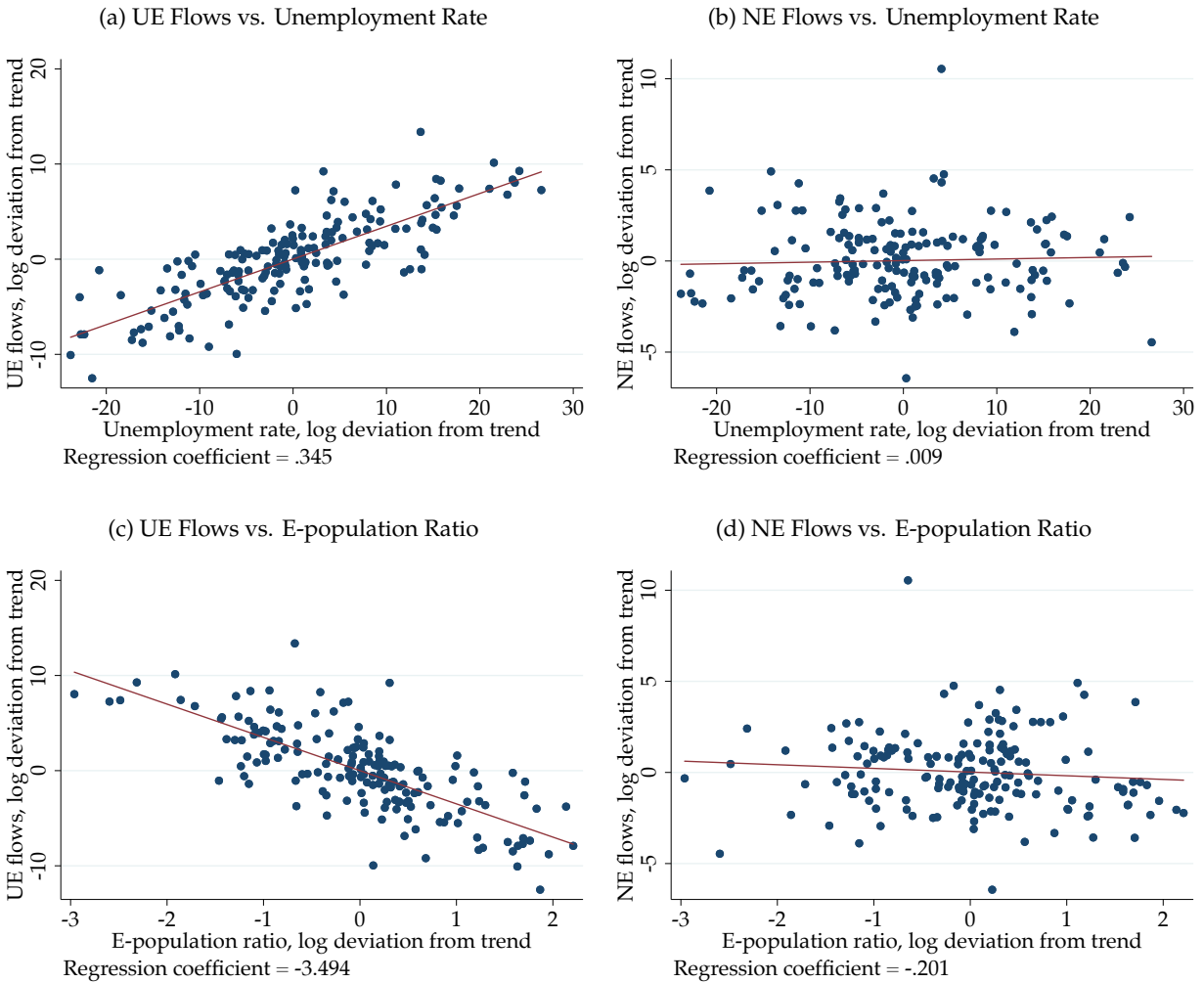
A.5 Evidence from OECD Countries

The countercyclicity of UE flows extends to many OECD countries. In Figure A10 Panel (a), we plot the elasticity of UE flows with respect to the unemployment rate for a set OECD countries, drawing on transition rates estimated in Elsby, Hobijn, and Şahin (2013) on the basis of labor force survey data and unemployment stocks.

As a validation check, we point out another perspective on the elasticity of UE flows with respect to the unemployment rate, repeated below for convenience

$$\varepsilon_{UE,u} = \frac{df/f}{du/u} + 1 = \left((1-u) \left[-1 + \frac{d\delta/\delta}{df/f} \right] \right)^{-1} + 1. \quad (\text{A8})$$

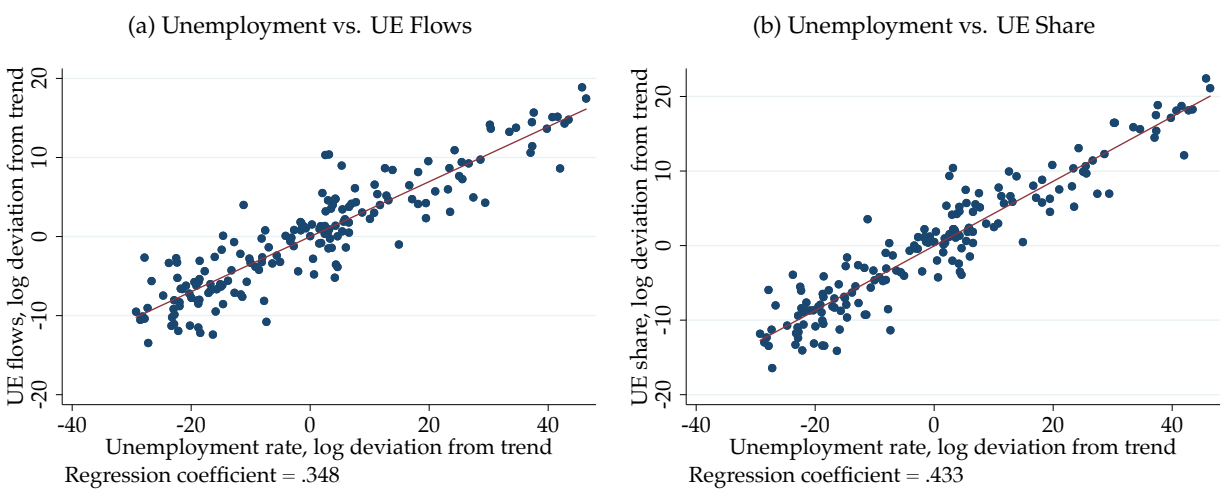
Figure A8: Cyclicity of New Hires: CPS Worker Flows



Notes: This figure is a complement to Figure A7. The figure plots different measures of new-hire flows into employment (UE or NE) against employment measures (unemployment rate or employment-population ratio). All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1,600. While our model relies of the share of new hires in employment rather than worker flows, this figure presents the cyclical behavior of nonemployment-to-employment flows, which are nearly acyclical, but importantly remain countercyclical as a share of (procyclical) employment, in turn presented in Figure A5. Source: CPS monthly files.

Building on the insight that the unemployment rate fluctuations implied by the job finding rate shift only is $du^f/u^f = -(1-u)df/f$. Fujita and Ramey (2009) show that the regression coefficient of du^f/u^f on du/u also represents the share of the variance in unemployment rate fluctuations due to fluctuations in the job finding rate (rather than in the job separation rate). The smaller this share, the more countercyclical the UE flows on average, since $\frac{dUE/UE}{du/u} = -\frac{1}{1-u} \frac{du^f/u^f}{du/u} + 1$. Drawing on cross-country differences in the OECD, we document the empirical validity of this theoretical property in Panel (b) of Figure A10, a scatterplot that shows a clear negative relationship between the elasticity against the contribution of job finding rate to unemployment fluctuations, the latter

Figure A9: The Countercyclicality of Unemployment-to-Employment Flows

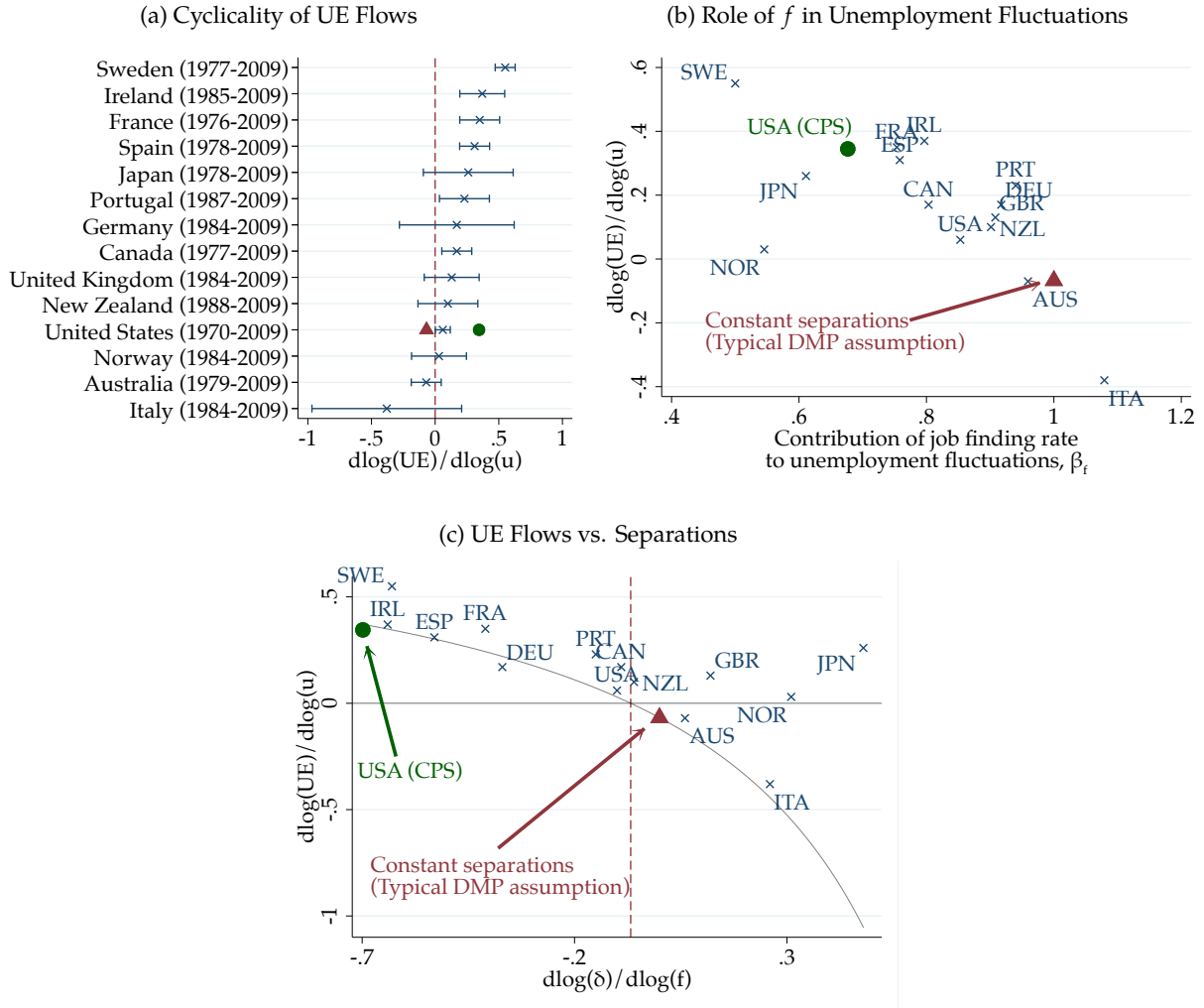


Notes: Panel (a) plots the log deviations in UE flows and log deviations in the unemployment rate from their respective trends. Panel (b) plots log deviations in UE share in employment against log deviations in the unemployment rate. All series are based on quarterly averages of monthly data. Detrended series are HP filtered with a smoothing parameter of 10^5 . Source: CPS monthly files.

computed in Elsby, Hobijn, and Şahin (2013). Since we apply steady-state approximations while Elsby, Hobijn, and Şahin (2013) point out that in many OECD countries dynamic expressions are appropriate, and since the unemployment rates are not homogeneous, this scatter plot does not trace out a perfectly straight line.

Finally, Panel (c) plots the UE flows-unemployment rate elasticity against the job finding-job separation rate elasticity in our sample of OECD countries, together with the theoretical relationship between the two as determined by Equation (A8). Broadly, the relationship between the two elasticities holds across countries (with the approximation error reflecting the assumptions of stationarity and having only two labor market states).

Figure A10: Cyclicity of UE Flows in the OECD

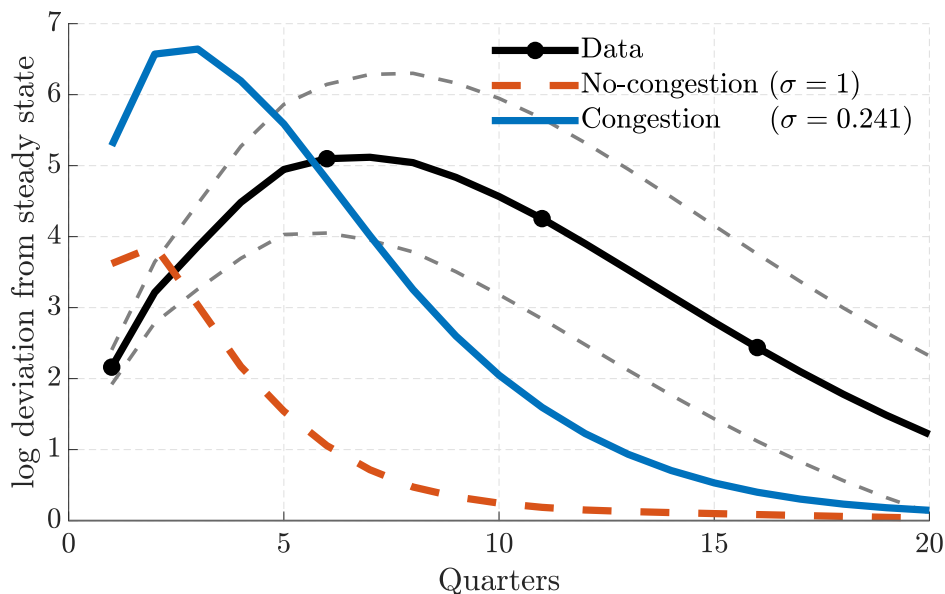


Notes: Panel (a) plots the elasticity of UE flows with respect to the unemployment rate in a set of OECD countries. Panel (b) plots these elasticities against the importance of job finding rate fluctuations in explaining the volatility in unemployment for each country. To compute the contribution of the job finding rate to unemployment fluctuations based on monthly CPS data (green dot), we calculate $\text{cov}(-(1 - \bar{u}_{ss})\hat{f}, u_{ss})/\text{var}(u_{ss})$, where u_{ss} is the steady-state approximation to the unemployment rate, \bar{u}_{ss} is its trend and \hat{f} is the cyclical component of (log) job finding rate (see Fujita and Ramey, 2009), such that $-(1 - \bar{u}_{ss})\hat{f}$ is the unemployment rate deviation due to the job finding rate only. For the DMP model without separation shocks, this share is one, and the elasticity on the y-axis is computed using formula (A8). Panel (c) plots the elasticity of UE flows with respect to the unemployment rate as well as the theoretical relationship between the two based on a steady state approximation. Source: Elsby, Hobijn, and Şahin (2013) and CPS monthly files.

B Further Details on Identification of Congestion

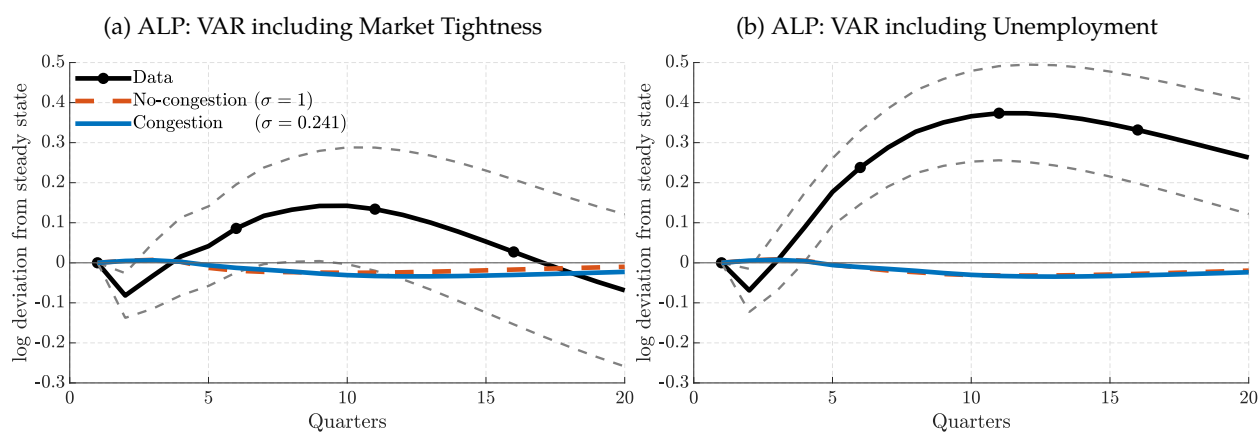
B.1 Additional Tables and Figures

Figure A11: IRF of Unemployment Rate to a Separation Rate Shock: Data and Models



Notes: The figure plots the empirical impulse response of unemployment rate to a separation shock (dashed lines are one standard deviation confidence bands), together with model responses. The red dashed line is the standard model with homogeneous workers ($\sigma = 1$). The blue solid line is our model under the preferred calibration ($\sigma = 0.241$).

Figure A12: IRF of Labor Productivity to a Separation Rate Shock: Data and Models



Notes: Panel (a) plots the impulse response of average labor productivity to a unit standard deviation job separation shock using the VAR model in Equation (7) with market tightness as the last variable. Panel (b) plots the same with unemployment rate as the last variable.

Table A3: Benchmarks: Elasticity of Substitution Parameters Estimated in the Literature

Paper	Production factors	Specification	Congestion Parameter
Mercan, Schoefer, Sedlacek (2021)	Workers indexed by periods since unemployment n_k	$Y = z(\sum_k \alpha_k n_k^\nu)^{1/\sigma}$	$\sigma = 0.241$
Krusel, Ohanian, Rios-Rull, Violante (2000)	High skill h and low skill l workers	$Y = Ak_s^\alpha (\mu l^\sigma + (1-\mu)(\lambda k_e^\rho + (1-\lambda)h^\rho)^{\frac{\sigma}{\rho}})$	$\sigma = 0.401$
Hagedorn, Manovski, Stetsenko (2016)	Capital structures k_s and capital equipments k_e		
Jeong, Kim, Manovskii (2015)	Worker experience E and labor L	$Y = A(L^\mu + \delta E^\mu)^{1/\mu}$	$\mu = -2.95$
Borjas (2003)	Capital K and labor L	$Q = (\lambda_k K^\nu + \lambda_l L^\nu)^{1/\nu}$	$\eta = 0.714$
	Worker with education i and experience j , L_{ij}	$L = (\sum_i \theta_i L_i^\rho)^{1/\rho}$, $L_i = (\sum_j \alpha_{ij} L_{ij}^\eta)^{1/\eta}$	$\rho = 0.231$
Autor, Katz, Kearney (2008)	College N_c and high school N_h equivalent workers	$Q = [\alpha(N_c)^\rho + (1-\alpha)(bN_h)^\rho]^{1/\rho}$	$\rho = 0.363$
Card, Lemieux (2001)	High school and college labor of different age groups H_j and C_j	$Y = (\theta_h H^\rho + \theta_c C^\rho)^{1/\rho}$ $H = \sum_j (\alpha_j H_j^\eta)^{1/\eta}$, $C = \sum_j (\beta_j C_j^\eta)^{1/\eta}$	$\rho = 0.091 - 0.375$ $\eta = 0.75 - 0.83$

Notes: This table presents a meta study of existing estimates of elasticities of substitution in CES aggregate production functions with different labor types (of various kinds), reformulated analogously to our congestion parameter σ .

B.2 Robustness of Identification of Separation Shocks using Time-Series Variation

The main text uses a three-variate VAR to identify exogenous separation shocks, which are crucial for quantifying our congestion mechanism. In particular, the response of labor market tightness to the separation shock, identified recursively using a Cholesky decomposition, is the key moment that pins down our preferred value of σ , which governs the extent of congestion.

More so than in cross-sectional studies, shocks other than labor productivity may be correlated with separation rate shifts in the aggregate time series (see Uhlig, 2005, for standard concerns with the VAR approach). After all, ALP is smooth and not very cyclical (see, e.g., Shimer, 2005; Mitman and Rabinovich, 2020; Galí and Van Rens, forthcoming). In fact, in canonical models of endogenous separations (Mortensen and Pissarides, 1994), the same surplus shock that drives hiring fluctuations, drives separations. At the same time, however, separation and job finding rates exhibit considerable independent variation (see, e.g., Shimer, 2012), and there exist theories of separation rate fluctuations without any connection to job surplus fluctuations (e.g., Golosov and Menzio, 2020). Similarly, reallocation shocks (Lilien, 1982) may shift new and old jobs' values in a directly affected sector, with the absorption of the freed-up labor to be done in a sector in which, e.g., TFP has not changed.

To address these concerns, this section assesses the role of omitted shocks in our estimated separation rate process. In particular, we study the leading drivers of business cycles in the macro literature: shocks to utilization-adjusted total factor productivity (Fernald, 2014), credit spreads, (Gilchrist and Zakrajšek, 2012), discount factors (Hall, 2017), uncertainty (Jurado, Ludvigson, and Ng, 2015), and monetary policy (Romer and Romer, 2004; Wieland and Yang, 2020). We find that these shocks have essentially no predictive power for the separation shocks identified by our VAR. Moreover, controlling for these shocks leaves the specific time-path of our separation shocks essentially unchanged.⁴ We conclude that the leading candidates of observable shocks are unlikely to confound our estimation of the congestion dynamics.

B.3 Data for Alternative Shocks

We now describe the data used for our analysis. The three-variate VAR is the same as in the main text, described in Section 2. The data for the other macroeconomic shocks are described below.

Total Factor Productivity Shocks. We take the utilization-adjusted quarterly measure of total factor productivity (`dtfp_util`) from Fernald (2014). The sample period for this shock is 1976Q1 – 2019Q4.

Financial Shocks. We use the “Gilchrist-Zakrajšek” credit spread as measured in Gilchrist and Zakrajšek (2012). The sample covers 1976Q1 – 2010Q3.

⁴An alternative route would be to include those shocks in the empirical VAR. Since our theoretical model will not feature those shocks, we do not pursue this route. We suspect that our results will be similar, since the VAR, intuitively, captures the residual variation of labor market tightness with separation shocks.

Discount Factor Shocks. We use the discount factor shocks estimated by Hall (2017), using the Shiller price index. The sample period is 1976Q1 – 2015Q2.

Uncertainty Shocks. We use the one-quarter-ahead macroeconomic uncertainty shocks estimated by Jurado, Ludvigson, and Ng (2015). The sample period for this shock is 1976Q1 – 2019Q4.

Monetary Policy Shocks. We use the monetary policy shocks proposed by Romer and Romer (2004) and as updated by Wieland and Yang (2020). The sample period for this shock is 1976Q1 – 2007Q4.⁵

B.4 Separation Shocks and Other Macroeconomic Disturbances

To ascertain whether our estimated separation shocks are not simply reflecting effects of omitted variables, we regress them on the range of macroeconomic shocks described above. Specifically, we estimate

$$\delta_t = \alpha_j + \sum_{s=0}^p \beta_{j,s} x_{j,t-s} + \eta_{j,t}, \quad (\text{A9})$$

where $x_{j,t}$ indicates a structural shock in period t , where j denotes one of the five structural shocks (TFP, financial, discount factor, uncertainty and monetary policy). We choose $p = 4$, thereby considering the contemporaneous impact of the structural shocks as well as up to four of their quarterly lags.

In addition to estimating the individual impact of each of the macroeconomic shocks, we also consider their joint effect by estimating

$$\delta_t = \tilde{\alpha}_j + \sum_{j=1}^5 \sum_{s=0}^p \tilde{\beta}_{j,s} x_{j,t-s} + \eta_t. \quad (\text{A10})$$

In all the above cases, we always estimate the regressions on the maximum sample size allowed by the data.

Table A4 presents the adjusted R^2 from each of the specifications above. The results suggest that the separation shocks identified by our three-variate VAR are in fact *not* driven by other (omitted) structural shocks that are independently identified outside of our VAR. The highest explanatory power is obtained by considering discount factor shocks, but even there the adjusted R-square is only 1.6%.

Figure A13 shows how the separation shocks estimated in the main text change when controlling for all of the above macroeconomic shocks using the regression model in Equation (A10). The figure reveals that the estimated shocks are largely unchanged, as suggested by the slightly negative R^2 in Table A4.

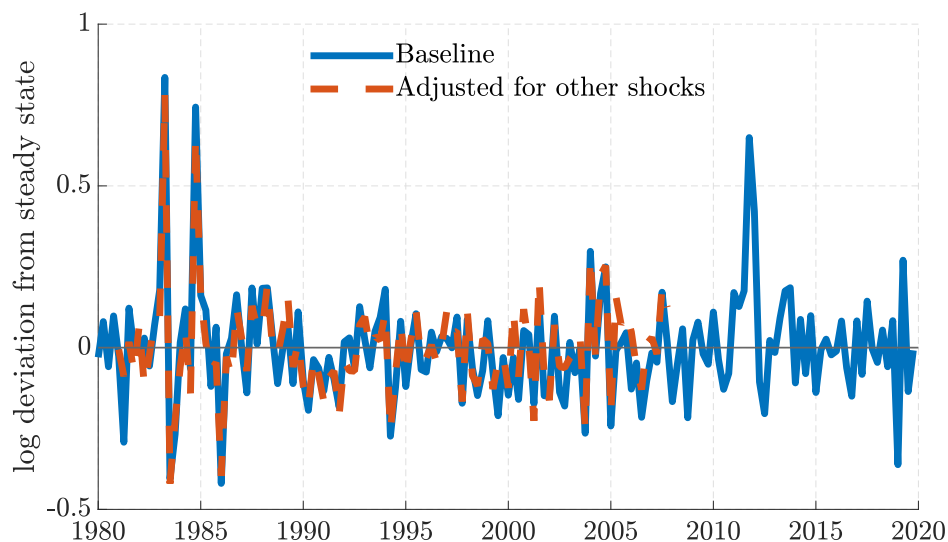
⁵An alternative approach is to identify monetary policy shocks using high-frequency identification as in, e.g., Gürkaynak et al. (2005); Gorodnichenko and Weber (2016); Gertler and Karadi (2015). However, these shock series cover a considerably shorter sample period.

Table A4: Separation Shocks and Other Disturbances: Adjusted R-squared

	TFP	Financial	Discounts	Uncertainty	Monetary Policy	Joint
R^2	-0.006	0.004	0.016	-0.007	-0.018	-0.096
# of obs.	156	119	138	156	108	108
# of coefs.	6	6	6	6	6	26

Notes: The top row reports the adjusted R-square from the individual regressions (A9) for the five different macroeconomic shocks and the “joint” regression in Equation (A10). “TFP” is the utilization-adjusted total factor productivity (Fernald, 2014), “financial” is the “Gilchrist-Zakrajšek” credit spread (Gilchrist and Zakrajšek, 2012), “discounts” is the discount factor shock based on the Shiller price index (Hall, 2017), “uncertainty” is the one-quarter-ahead macroeconomic uncertainty (Jurado, Ludvigson, and Ng, 2015) and “monetary policy” is taken from Wieland and Yang (2020). The second and third rows report, respectively, the number of observations and estimated parameters in each regression.

Figure A13: Separation Shocks: Baseline and Adjusted for Identified Shocks



Notes: The figure shows the baseline separation shocks estimated in Section 2 and those shocks “adjusted for other disturbances” using the regression model in Equation (A10), where the plotted series is given by η_t .

C A Generalization of the Baseline Model: Types vs. Inputs

The baseline model in the main text assumes that every worker type k is a different input in production, i.e., $Y = z \left(\sum_{k=1}^K \alpha_k n_k^\sigma \right)^{\frac{1}{\sigma}}$. In this appendix we generalize this setup by allowing for subsets of worker types $i \subset \mathcal{K}$ to be perfectly substitutable in production. That is, different types k are not necessarily separate worker types as inputs into production, i . Instead, an input type $i \in \mathcal{I} = \{1, \dots, I\}$ is defined by a set of worker types $\Omega_i \subset \mathcal{K}$ which are mutually exclusive, i.e., $\cap_i \Omega_i = \emptyset$. The production function in this setting is given by $Y = z \left(\sum_i \alpha_i n_i^\sigma \right)^{\frac{1}{\sigma}}$.

This setup of worker heterogeneity nests multiple cases. For example, if $I = 1$, then $\Omega_1 = \mathcal{K}$ and all worker types constitute one input type (homogeneous workers). Types do not matter for production, so that this case boils down to the standard DMP model with a redundant worker type evolution in the background. Another setup has low- and high-skilled workers, where the former become the latter after, e.g., three years of employment. In a quarterly calibration, this setup would be given by assuming $I = 2$ with $\Omega_1 = \{1, \dots, 12\}$ and $\Omega_2 = \{13, \dots, K\}$. As a final example, each worker type is a separate input type (as in the main text), in which case $I = K$, and $\Omega_i = \{i\}$ for $i = 1, \dots, K$.

The retailer buys $\{n_i\}_{i=1}^I$ units of output in a perfectly competitive market. This implies that the prices for these goods satisfy the static first order conditions:

$$p_i = \alpha_i n_i^{\sigma-1} \frac{Y}{\sum_j \alpha_j n_j^\sigma} = \alpha_i s_i^{\sigma-1} \frac{1}{\sum_j \alpha_j s_j^\sigma} \frac{Y}{N}, \quad (\text{A11})$$

where $s_i = n_i/N$ denotes the share of type- i workers in production, and $N = \sum_i n_i$ is aggregate employment.

The worker and firm values now reflect the fact that worker types themselves are not imperfect substitutes in production, but only through their position in the production sets $i(k)$. The model equations differ only in that worker heterogeneity is now indexed by $i(k)$, rather than k .

D Solution Method

This appendix provides details of the solution and estimation methods used in the paper. We begin by describing the computation of the steady state, which includes the distribution of worker types among the employed and unemployed. We then lay out the solution method for the dynamic model and for its estimation.

D.1 Steady State

Given our parameterization, in particular the matching of the steady state job finding and separation rates, and our assumption that all unemployed fall to $k = 1$, it is possible to compute the implied distribution of worker types without solving for the rest of the model. Specifically, the steady state distribution of employment across worker types and steady state unemployment can be solved from the following set of equations:

$$\begin{aligned} e_1 &= f u, \\ e_{k+1} &= e_k(1 - \delta) \quad \text{for } k = 1, \dots, K - 1, \\ u &= (1 - f)u + \delta \sum_k e_k. \end{aligned}$$

In addition, under our calibration ensuring that $p_k = 1$ for all k in steady state, it is possible to compute the steady state surplus values for each type. This result, in turn, also pins down the steady state value of labor market tightness via the free-entry condition in Equation (20). Finally, using the steady state distribution of employment levels, and again the assumption that $p_k = 1$ for all k in steady state, we can calculate the implied productivity weights α_k via

$$1 = p_k = a_k s_k^{\sigma-1} \frac{1}{\sum_{l=1}^K \alpha_l s_l^\sigma} \frac{Y}{N},$$

where $s_k = e_k / (\sum_{l=1}^K e_l)$, and where we normalize average labor productivity $Y/N = 1$.

D.2 Solution and Estimation with Aggregate Uncertainty

Our model features heterogeneity in worker types and two aggregate sources of uncertainty, z and δ . The employment distribution gives another set of endogenous state variables. The distribution is, however, described without approximation error by the masses of workers of each of the K types. Transitions between these types shown in Equation (10), which depend on the job finding and separation rates, describe the distributional movements over time.

Therefore, there is no need to revert to iterative procedures, as the law of motion for the distribution is known a priori. We solve the model using first order perturbation around its stationary steady state (i.e., including the employment distribution). The large number of state

variables (the two aggregate shocks, the distribution of employment shares and the unemployment rate) do not impede the speed of the solution method as perturbation is not prone to the curse of dimensionality.

To compute business cycle statistics, we simulate the model 100 times for 176 quarters (the length of our empirical sample). For each simulation, we detrend the logarithms of all the variables using the HP filter with a smoothing parameter of 1,600. The reported statistics are then averages over the 100 simulations. This also applies to impulse responses, which are averages of the estimated VARs over the 100 simulations.

D.3 The Kalman Filter

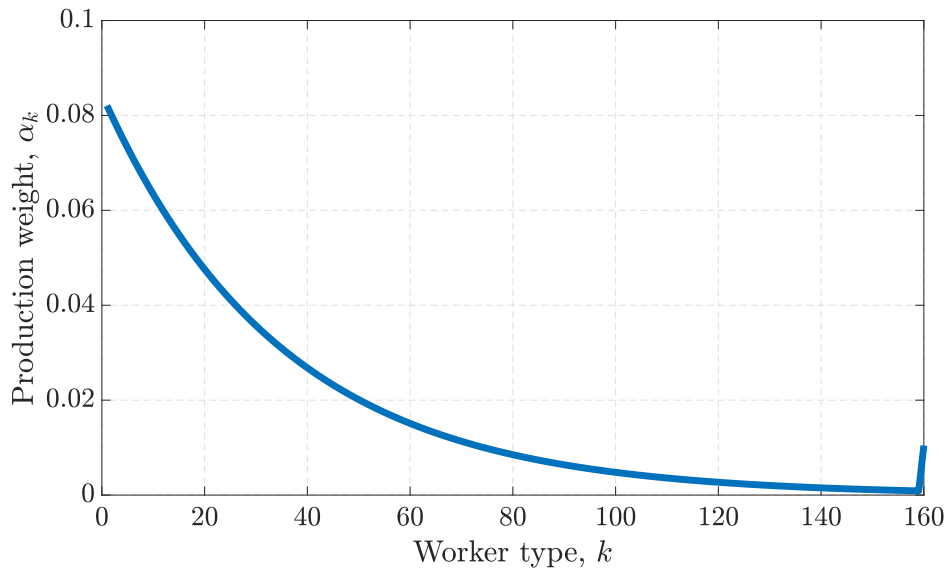
In addition, the linear nature of our solution allows us to estimate the model using the Kalman filter. Specifically, in Section 5.4 we use data on average labor productivity and the share of newly hired workers in employment to estimate the time path of the two aggregate shocks consistent with these two time series and our parameterization. The model structure then implies a particular time path for all model variables. We use this property in Section 5.4 to calculate the contribution of congestion unemployment to the variation in observed unemployment fluctuations. Figure A21 shows the time paths of other labor market variables implied by our estimation.

E Details of the Baseline Parameterization: Homogeneous Steady State Marginal Products Across Types

The main text describes the parameterization of the model, including that of the production weights α_k for different worker types. These are set such that the respective marginal products, p_k , are equal to 1 for all k . Hence, all worker types have the same (fundamental) surplus in steady state.

Figure A14 visualizes the calibrated values of the relative productivities. Their pattern mimics that of employment shares. Relatively abundant types, such as worker type $k = 1$, would be characterized by a lower marginal product unless its abundance is offset by a higher relative productivity weight α_1 . The spike at $k = K$ is due to the fact that this type is an absorbing state and therefore employment in this type is somewhat higher than in $k = K - 1$.

Figure A14: Relative Worker Productivities in the Congestion Model



Notes: The figure plots the relative weights in production, α_k , in the congestion model with $\sigma = 0.241$. The spike at $k = K (= 160)$ reflects the fact that it is an absorbing state.

F Alternative Calibration: Small Surplus/“High b ”

It is well understood that low fundamental surplus values help amplify the effects of productivity shocks and generate realistic unemployment fluctuations (see e.g., Ljungqvist and Sargent, 2017; Hagedorn and Manovskii, 2008). In this section, we consider an alternative calibration without congestion ($\sigma = 1$) with low surplus.

We calibrate most of our parameters as in the main text, except for the flow value of unemployment b , which is set such that the model matches the volatility of labor market tightness. We consider a version with and without separation shocks. The implied value of b is 0.96 in the case without separation shocks.

Results are presented in Table A5. While the model without separation shocks matches—by construction—the volatility of labor market tightness, it fails on the cyclicity of UE flows, for the same reasons as discussed in Section 2: separation shocks are necessary to match the countercyclical nature of UE flows. In the case with separation shocks, the model matches well the volatility of essentially all labor market variables. In addition, the model now also matches the countercyclicity of UE flows, albeit to a lesser extent than in the data. However, it grossly fails in the response of labor market tightness to a separation shock, as the standard model with separation rate shock discussed in the main text.

Figure A15 shows the empirical response of labor market tightness to a separation shock, with that of the model without congestion but with a low fundamental surplus and separation shocks. As in the standard model without congestion, there is essentially no response of labor market tightness to a separation shock. This key result does not change with a low fundamental surplus.

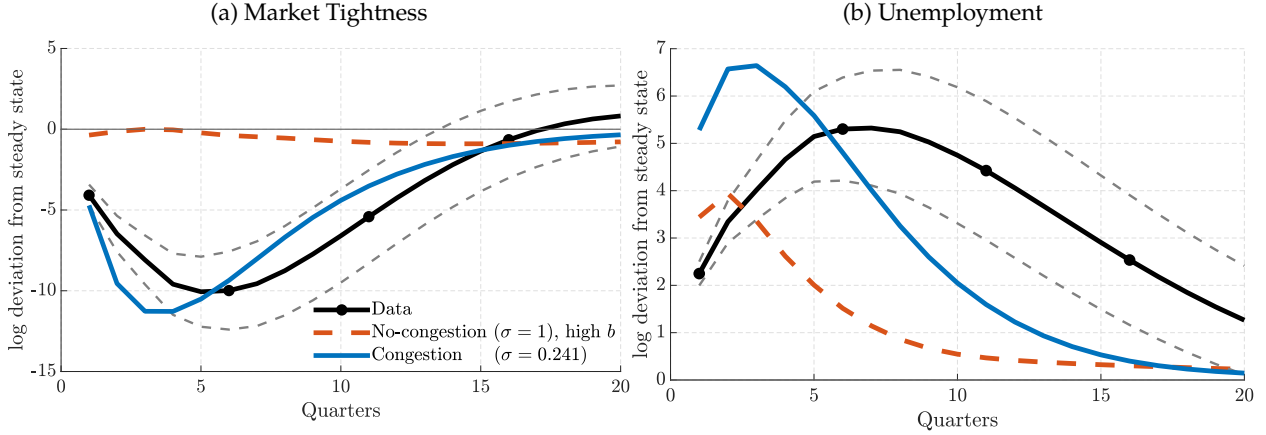
Steady State Elasticities. To understand this result further, we conduct a version of the analysis in Ljungqvist and Sargent (2017), but this time for separation shocks. In order to see whether separations have a sizable impact on hiring, we derive the elasticity of labor market tightness with respect to separations. Following Ljungqvist and Sargent (2017), we cast our model in continuous time in which case the hiring condition can be written as

$$r + \delta = \frac{(z - b)(1 - \phi)q(\theta)}{\kappa} - \phi f(\theta), \quad (\text{A12})$$

where r is the interest rate such that $\beta = 1/(1 + r)$. Taking z as given and totally differentiating Equation (A12) with respect to δ and θ gives

$$\begin{aligned} d\delta &= \frac{(z - b)(1 - \phi)q'(\theta)}{\kappa} d\theta - \phi f'(\theta) d\theta \\ &= -[\mu(r + \delta) + \phi f(\theta)] \frac{d\theta}{\theta}. \end{aligned} \quad (\text{A13})$$

Figure A15: Impulse Responses to a Separation Shock: No-Congestion, Low-Surplus Model



Notes: The figure plots the impulse responses of labor market tightness and unemployment rate to a separation shock in the data and model, which is calibrated under a low fundamental surplus (e.g., Hagedorn and Manovskii, 2008) and includes countercyclical separation shocks.

Rearranging the above, we can then write the elasticity of θ with respect to δ as

$$\epsilon_{\theta, \delta} = \frac{d\theta/\theta}{d\delta/\delta} = -\frac{\delta}{\mu(r + \delta) + \phi f(\theta)} = -\Upsilon^{\text{Nash}} \frac{\delta}{r + \delta + \phi f(\theta)}, \quad (\text{A14})$$

where $\Upsilon^{\text{Nash}} = \frac{r + \delta + \phi f(\theta)}{\mu(r + \delta) + \phi f(\theta)}$ is the scaling factor, which multiplies the fundamental surplus, derived in Ljungqvist and Sargent (2017). As discussed in Ljungqvist and Sargent (2017), reasonable calibrations of the standard search and matching model results in $\Upsilon^{\text{Nash}} \approx 1$. Moreover, these calibrations also result in the denominator in Equation (A14) being roughly equal to one half. In conclusion, the standard model features labor market tightness which is largely insensitive to separation shocks, with an elasticity of around -2δ . Moreover, this elasticity is *independent* of the fundamental surplus. This is precisely the reason why even a calibration with a low fundamental surplus cannot replicate the empirical response of labor market tightness to separation shocks.

Table A5: Business Cycle Properties: No-Congestion, Low-Surplus Model

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
<i>Panel A: Low Fundamental Surplus Model Without Separation Shocks</i>							
Standard deviation	0.010	0.064	0	0.052	0.199	0.230	0.049
Autocorrelation	0.706	0.706	0	0.844	0.596	0.706	0.311
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.999	1					
δ	0	0	1				
<i>u</i>	-0.647	-0.648	0	1			
<i>v</i>	0.980	0.981	0	-0.486	1		
θ	0.999	1.000	0	-0.648	0.981	1	
<i>UE/E</i>	0.476	0.476	0	-0.270	0.477	0.476	1
<i>Panel B: Low Fundamental Surplus Model With Separation Shocks</i>							
Standard deviation	0.010	0.064	0.082	0.090	0.177	0.227	0.068
Autocorrelation	0.691	0.689	0.560	0.825	0.558	0.689	0.623
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.999	1					
δ	-0.413	-0.430	1				
<i>u</i>	-0.674	-0.684	0.699	1			
<i>v</i>	0.933	0.929	-0.197	-0.368	1		
θ	0.999	1.000	-0.430	-0.684	0.929	1	
<i>UE/E</i>	0.005	-0.001	0.266	0.455	0.229	-0.001	1

Notes: *ALP*, *f*, δ , *u*, θ and *UE/E* indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, and share of new hires in employment. Panel A reports values from the model with a constant separation rate, Panel B reports the same for the model with countercyclical job separation shocks. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1,600.

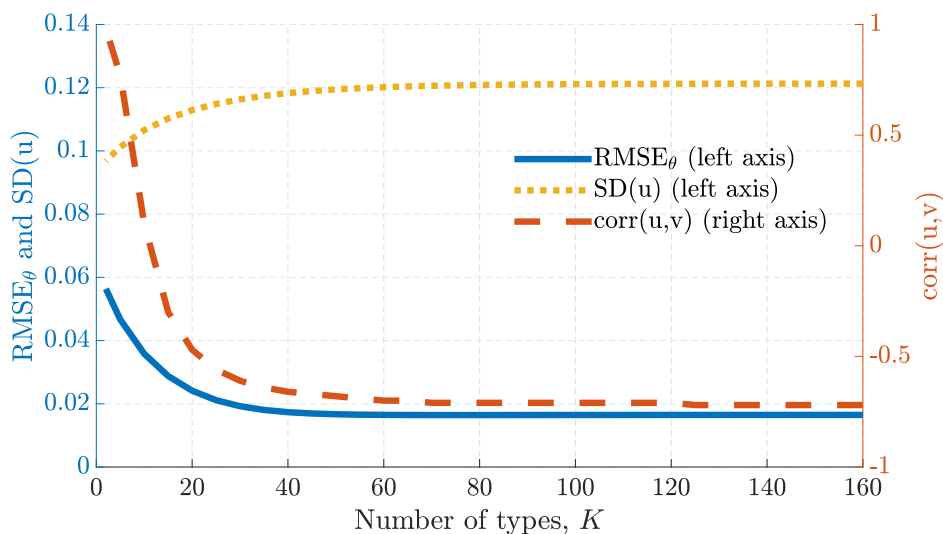
G Robustness of Model Performance to the Number of Worker Types

In the baseline model, we assume a maximum of $K = 160$ worker types. This means that workers gradually progress through different types while employed and remain at the highest rung only after 40 years of continuous employment in a quarterly calibration. In this section, we show that our results are robust to considerably reducing K .

Specifically, we solve and simulate our model for $K = 2, 3, \dots, 160$ and plot the corresponding standard deviation of unemployment, slope of the Beveridge curve ($\text{corr}(u, v)$) and the root mean squared error between the model and empirical impulse responses of labor market tightness to a job separation shock for each value of K . While the former two are key model outcomes, the latter is a measure used to parameterize the congestion parameter, σ . Note that in this exercise we do not recalibrate σ (but keep it at the value in the baseline model with $K = 160$), nor the size of the shocks.⁶

Figure A16 shows the three measures as a function of the number of types K on the horizontal axis. Our baseline results are obtained on the far right of the figure, at $K = 160$. As the figure makes clear, our results are essentially unaffected even if the number of types K is reduced from 160 to about 30, i.e. even if workers “grow out” of their cohort after about 7 years, rather than 40.

Figure A16: Volatility of unemployment, slope of the Beveridge curve and model fit vs K



Notes: The figure plots the standard deviation of unemployment, the correlation between unemployment and vacancies (i.e., the slope of the Beveridge curve) and the distance between model and empirical IRFs of labor market tightness to an innovation in the job separation rate for each of the of maximum number of worker types K .

Reducing the number of types further below 30 leads to a quick deterioration of the model performance. With the extreme case of only $K = 2$ types (far left values), we effectively retain the “standard” search and matching model with separation shocks, which completely fails in

⁶We do, however, recalibrate the production weights, α_k , such that the marginal product of each worker type p_k still remains to be equal to 1 as in the baseline.

generating a strongly negatively sloped Beveridge curve. In fact, in the case of $K = 2$ worker types, the correlation between unemployment and vacancies is almost 1. Moreover, this model fails to generate congestion as measured by the effect of labor market tightness to separation rate shocks, as indicated by the RMSE curve. Hence, the region where K yields low levels of amplification is exactly the region where the model starts to fail on other margins too; this is due to the absence of congestion, and may require recalibrating σ (which, in turn, would recover similar amplification even with lower K). We here illustrate that even with the original σ , a wide range of K leaves our results unchanged.

To understand these patterns, it is useful to express the surplus of newly hired workers as the discounted present value of *future* productivity levels (net of outside options, b). In particular, in our model the surplus of a worker of type k can be expressed as (see Equation (23)):

$$S_{k,t} = p_{k,t} - b + \beta \mathbb{E}_t(1 - \delta_{t+1})S_{k+1,t+1} - \beta \mathbb{E}_t(1 - \delta_{t+1})f(\theta_t)\phi S_{1,t+1}. \quad (\text{A15})$$

Focusing on the steady state and iterating the above equation forward (in terms of worker cohorts, k), we can express the surplus of a newly hired worker as

$$S_1 = \sum_{k=1}^{\infty} (\beta(1 - \delta))^{k-1} (p_k - b) - \sum_{k=1}^{\infty} (\beta(1 - \delta))^k f(\theta)\phi S_1. \quad (\text{A16})$$

As in the standard search and matching model, the surplus of newly hired workers determines hiring decisions and, therefore, the pattern of vacancies. In our baseline model, a separation shock incipiently increases unemployment and subsequently the mass of new hires in aggregate production, which leads to a reduction in $p_{1,t}$ through our congestion mechanism. In addition, as we explain in the main text, workers cannot escape the curse of an abundant cohort because all newly hired workers who remain in employment become $k = 2$ type workers in $t + 1$, $k = 3$ type workers in period $t + 2$ and so on. Only after reaching the final type $k = K$ do cohorts “mix.” Therefore, our model features strong cohort effects, which further reduce the incentives to hire because workers are expected to have depressed marginal products in the future as they cannot escape the abundant hiring cohort for up to $K = 160$ quarters.

As we reduce K , workers can escape their abundant recession cohort sooner. Whether or not the reduction in K affects the fluctuations in the total surplus of newly hired workers depends on a number of issues: how quickly workers escape their cohort (k), how plentiful is the absorbing type (K), how strongly its marginal product changes, the strength of discounting, and also the persistence of the hiring boom (as workers reaching the K state will benefit from their relative scarcity compared to younger cohorts). This is a quantitative question and it turns out that under our baseline calibration, the results are largely unchanged for K between about 30 and 160.

To understand the horse-race between these effects further, consider the extreme case of $K = 2$ and, as the first case, a persistent hiring burst. A separation shock reduces the employment stocks of both types, but because of full type loss in unemployment, only $k = 1$ types become abundant

at the time of hiring. However, $k = 2$ types remain scarce owing to the hiring burst into the $k = 1$ type. Therefore, the present-value surplus of newly hired workers depends on a depressed $p_{1,t}$ in the first period and an increased $p_{2,t+1}$ in the next period and thereafter, as the economy keeps absorbing new hires. Given that $k = 2$ is the absorbing highest worker type rung, a worker spends the vast majority of their employment in that type and the surplus of newly hired workers is dominated by the increase in p_2 . As a result, the surplus of new hires falls by less than in a model with higher K . As the second case, now consider $K = 2$ with a transitory hiring burst (lasting only the first period). Then, even with $K = 2$, congestion effects are larger as hiring returns to normal once the $K = 2$ type enters the $k = K = 2$ type, which is relatively abundant once the cohort enters it.

Finally, the effects described above depend on the relative amounts of time spent in abundant and scarce worker cohorts, rather than the absolute number of types. For simplicity, we have assumed that types correspond to quarters, and that workers move up by a type each model period. Alternatively, one could assume that workers move types, e.g., once every five years, i.e. we could slow down the type upgrade process without increasing K . In such a case, even a relatively small number of types K may allow for strong cohort effects and recover our baseline results (especially when recalibrating σ as in the main text).

H Business Cycle Statistics Including Full Correlation Matrices

For compactness, Table 2 in the main text only reports correlations with unemployment. Here, we additionally report the tables with the full correlation matrices. The tables are ordered as the panels in Table 2.

Table A6: Business Cycle Properties in the Data

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
Standard deviation	0.010	0.053	0.067	0.103	0.126	0.229	0.067
Autocorrelation	0.746	0.871	0.773	0.934	0.926	0.936	0.836
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.042	1					
δ	-0.415	-0.715	1				
<i>u</i>	-0.112	-0.931	0.848	1			
<i>v</i>	0.309	0.874	-0.869	-0.934	1		
θ	0.223	0.917	-0.874	-0.980	0.986	1	
<i>UE/E</i>	0.173	-0.722	0.567	0.833	-0.711	-0.783	1

Notes: *ALP*, *f*, δ , *u*, θ and *UE/E* indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and the share of new hires in employment. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

Table A7: Business Cycle Properties in the No-Congestion Model without Separation Shocks

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
Standard deviation	0.010	0.004	0	0.003	0.013	0.015	0.003
Autocorrelation	0.704	0.704	0	0.843	0.592	0.704	0.306
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	1.000	1					
δ	0	0	1				
<i>u</i>	-0.643	-0.643	0	1			
<i>v</i>	0.980	0.980	0	-0.481	1		
θ	1.000	1.000	0	-0.643	0.980	1	
<i>UE/E</i>	0.476	0.476	0	-0.272	0.476	0.476	1

Notes: *ALP*, *f*, δ , *u*, θ and *UE/E* indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and the share of new hires in employment. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

Table A8: Business Cycle Properties in the No-Congestion Model with Separation Shocks

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
Standard deviation	0.010	0.005	0.088	0.068	0.058	0.017	0.067
Autocorrelation	0.688	0.647	0.499	0.736	0.751	0.647	0.74
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.975	1					
δ	-0.441	-0.627	1				
<i>u</i>	-0.508	-0.665	0.916	1			
<i>v</i>	-0.306	-0.482	0.888	0.974	1		
θ	0.975	1.000	-0.627	-0.665	-0.482	1	
<i>UE/E</i>	-0.348	-0.402	0.413	0.739	0.747	-0.402	1

Notes: *ALP*, *f*, δ , *u*, θ and *UE/E* indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and the share of new hires in employment. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

Table A9: Business Cycle Properties in the Congestion Model—Baseline (Matching *UE/E*)

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>	p_1
Standard deviation	0.010	0.059	0.122	0.121	0.102	0.207	0.067	0.055
Autocorrelation	0.688	0.897	0.530	0.836	0.857	0.897	0.742	0.771
Correlation matrix								
<i>ALP</i>	1							
<i>f</i>	0.443	1						
δ	-0.410	-0.509	1					
<i>u</i>	-0.463	-0.924	0.743	1				
<i>v</i>	0.348	0.922	-0.157	-0.716	1			
θ	0.443	0.996	-0.514	-0.940	0.909	1		
<i>UE/E</i>	-0.337	-0.930	0.392	0.865	-0.876	-0.940	1	
p_1	0.490	0.952	-0.431	-0.862	0.900	0.949	-0.973	1

Notes: *ALP*, *f*, δ , *u*, θ , *UE/E* and p_1 indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, the share of new hires in employment and marginal labor product of new hires. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

Table A10: Business Cycle Properties in the Congestion Model—Robustness (Matching EU & Participation)

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>	p_1
Standard deviation	0.010	0.054	0.067	0.099	0.099	0.189	0.052	0.051
Autocorrelation	0.701	0.901	0.544	0.850	0.889	0.902	0.767	0.781
Correlation matrix								
<i>ALP</i>	1							
<i>f</i>	0.323	1						
δ	-0.419	-0.488	1					
<i>u</i>	-0.337	-0.941	0.693	1				
<i>v</i>	0.284	0.960	-0.241	-0.819	1			
θ	0.326	0.997	-0.491	-0.954	0.952	1		
<i>UE/E</i>	-0.240	-0.938	0.390	0.890	-0.913	-0.946	1	
p_1	0.414	0.950	-0.443	-0.882	0.926	0.948	-0.973	1

Notes: *ALP*, *f*, δ , *u*, θ , *UE/E* and p_1 indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, the share of new hires in employment and marginal labor product of new hires. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

I Alternative Calibration: Accounting for Labor Force Participation and Matching EU Flows

The baseline model calibrates separation shocks such that the model matches the observed fluctuations in the share of new hires in employment, UE/E , which are key to our congestion mechanism. However, as a result, the baseline model overpredicts the volatility of employment-to-unemployment (EU) flows, by overpredicting the volatility of EU separation rate δ .

In this appendix, we show that this inability to match both realistic new-hire employment shares and EU separations is primarily due to the missing non-participation margin in our two-state framework. We choose our two-state labor market framework for convenience and its direct comparability with canonical models in the literature (see, e.g., Shimer, 2005; Pissarides, 2009). However, two labor-market states mean that our framework attributes any flows into and out of non-participation (out of the labor force; OLF, or “O”, as we denoted nonemployment, comprising out of the labor and unemployment by “N”) to flows between employment and unemployment. This problem is common to all two-state models. See Elsby, Hobijn, and Şahin (2015a) for the importance of the nonparticipation margin over the business cycle.

This quantitative extension still generates unemployment fluctuations that are 96% as volatile as in the data, and the Beveridge curve correlation of -0.819 , indicating that the success of the model is robust to alternative specification of worker flows. Therefore, our preferred specification remains the simple two-state model for convenience and its direct comparability with canonical models in this active literature (see, e.g., Shimer, 2005; Pissarides, 2009; Hagedorn and Manovskii, 2008; Ljungqvist and Sargent, 2017).

I.1 Clarifying the Problem

One consequence of the omitted third state and transitions into and out of it is that the law of motion for unemployment—which holds in the model at all times—does not hold for the empirical measures of f , δ and u :

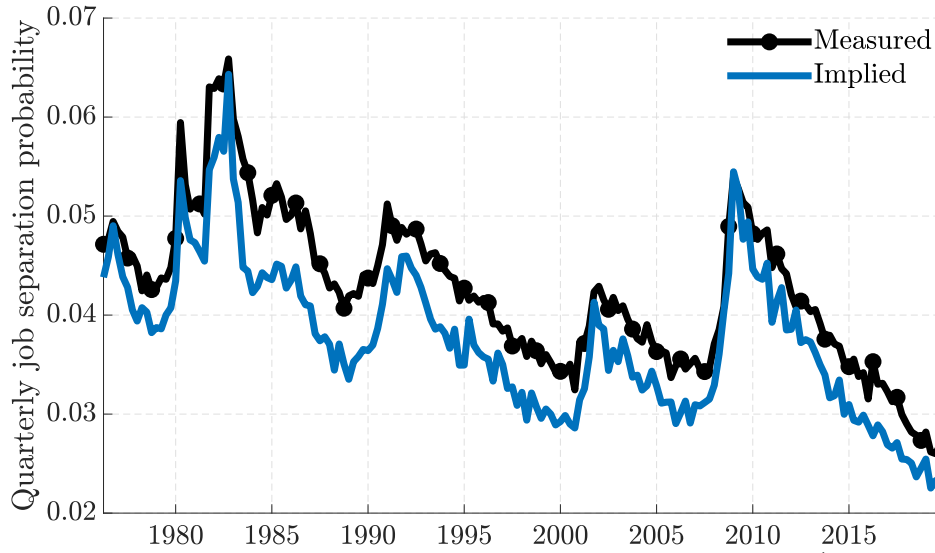
$$u_{t+1} = (1 - f_t)u_t + \delta_{t+1}(1 - u_t). \quad (\text{A17})$$

However, it is possible to compute an *implied* empirical measure of EU separation rates $\delta_{t+1}^{\text{imp}}$ consistent with the two-state law of motion of unemployment given by the above equation and the measured unemployment and job finding rates in the data.⁷ Specifically, we compute this implied process as

$$\delta_{t+1}^{\text{imp}} = \frac{u_{t+1} - (1 - f_t)u_t}{1 - u_t}. \quad (\text{A18})$$

⁷This procedure resembles that in Shimer (2005), who backs out the job finding rate using the law of motion for unemployment and a proxy for EU flows using short-term unemployment. In our case, the procedure is reversed, with the EU flows being backed out from the law of motion for unemployment given a measure of the job finding rate.

Figure A17: Separation Rate: Measured and Implied



Note: “Measured” δ refers to the empirical time series in the main text. “Implied” refers to δ^{imp} described above, based on the law of motion for unemployment, the unemployment rate and the job finding rate.

In words, this implied separation rate captures the two-state separation rate process that would, when feeding in the empirical job finding rate and the unemployment rate, exactly predict the empirical level of the next period unemployment rate.

Comparing the implied EU separation rate δ^{imp} with the actual EU separation rate δ permits a useful diagnostic: whenever δ^{imp} exceeds δ , it must be that out-of-steady-state transitions between OLF and E or U occurred that, on net, lowered empirical employment or raised unemployment by more than accounted for by EU transitions (δ) and UE transitions (f)—where these have been constructed on the basis of panel data measuring the transitions of workers between U and E states, i.e. measured EU and UE flows.

Figure A17 shows the time series of the measured and implied separation rates, i.e., δ and δ^{imp} . The implied separation rate is more volatile and less persistent compared to the measured one. This comparison highlights the tension between a two-state model of the labor market and directly measured flows in the data.

Below, we recalibrate the baseline model to account for the discrepancy described above. Subject to a recalibration of our key parameter, σ , the extended model delivers essentially the same quantitative results while, at the same time, matching the observed variation in EU flows.

I.2 Introducing Flows Into and Out of Non-Participation

We now present an alternative model that quantitatively accounts for the presence of flows into and out of non-participation. To nevertheless retain the logic of our two-state model, we introduce *exogenous* net changes in the number of unemployed. In particular, the law of motion for the mass

of unemployed of type $k = 1$ is given by

$$u_{1,t} = (1 - f(\theta_{t-1}))u_{1,t-1} + \delta_t \sum_{k=1}^K e_{k,t} + OU_t, \quad (\text{A19})$$

where we have retained our assumption that all separated workers fall to the bottom of the ladder and become type $k = 1$. The new feature, compared to the baseline model, is the presence of OU_t flows, which reflects the possibility of (exogenous) changes in the unemployment pool proxying for flows into and out of OLF. Specifically, we assume that OU fluctuates according to the following process:

$$OU_t = \rho_{OU}OU_{t-1} + \epsilon_{OU,t}, \quad (\text{A20})$$

where $\rho_{OU} \in (-1, 1)$ is a persistence parameter and $\epsilon_{OU,t} \sim N(0, \sigma_{OU}^2)$ are random shocks. This extension does not change the steady state of our model as OU flows are assumed to have zero mean.

I.3 Parameterized Model with Empirical δ Process and OU Flows

We parameterize the extended model in exactly the same way as the baseline model, except that instead of targeting UE/E flows, we directly parameterize the δ process to match the cyclical pattern of EU flows in the data.

In addition, we set ρ_{OU} and σ_{OU} to match the persistence and volatility of OU flows as a share of the labor force observed in the data, constructed as described in Appendix A.1.⁸ In addition, we allow for a correlation between ϵ_δ and ϵ_{OU} to match the observed correlation between $OU/(E + U)$ and the unemployment rate, which is 0.72.

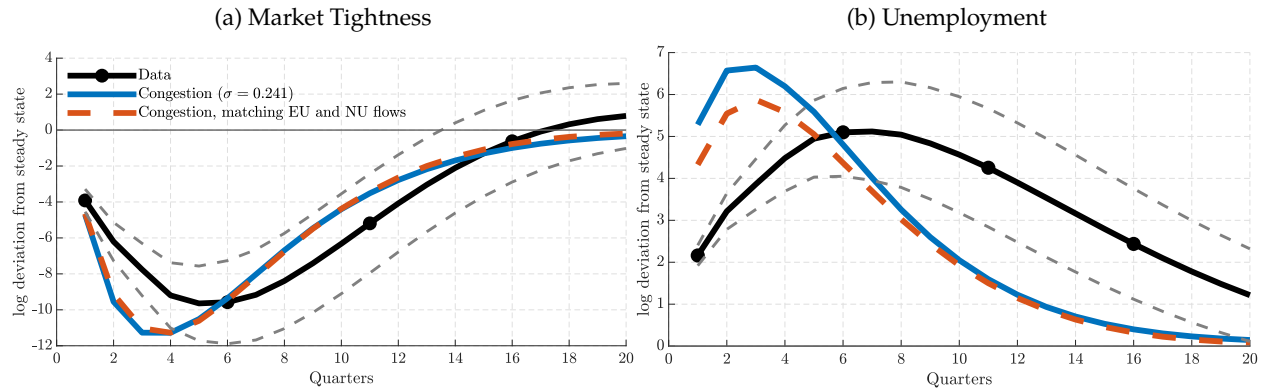
Table A10 in Appendix H, and Panel E of Table 2 in the main text, show the business cycle statistics of the extended model. This model matches not only the volatility of average labor productivity, but now also that of EU flows—specifically, the δ process now has the same volatility as in the data (although we miss some of its persistence). Moreover, the extended model still delivers a large amount of amplification of shocks. Specifically, the volatility of unemployment is 96% that of the data and the Beveridge curve has a healthy correlation of -0.819 . Since we no longer target the UE/E fluctuations, they are now somewhat less volatile than in the data. But, the calibrated separation shocks together with the additional OU flows result in unemployment-to-employment flows being relatively close to what they are in the data.

In order to match the impulse response of labor market tightness to separation shocks, the extended model requires a σ of 0.08. Under this calibration, however, the extended model delivers essentially identical dynamics as the baseline model, as shown in Figure A18.

To conclude, the baseline model refined to match the volatility of the empirical separation rate process and extended for the possibility of (exogenous) flows into and out of non-participation

⁸Because of the assumed zero mean in the model, we match the persistence and volatility in levels, rather than logs. The average ratio $OU/(U + E)$ is 1.4%, with persistence of 0.57 and standard deviation of 0.001.

Figure A18: Impulse Responses to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows



Notes: The figure plots the impulse responses of labor market tightness and unemployment rate to a separation shock in the data and model, which is calibrated to match the business cycle patterns of EU flows.

parameterized to match those observed in the data, has essentially identical amplification properties regarding labor market tightness and unemployment as the baseline model presented in the main text. While we choose to retain the standard two-state labor market model as our main specification, we conjecture that an explicit modelling of an endogenous non-participation choice would yield very similar results (provided that such a hypothetical model succeeds in matching the UE flows and congestion dynamics). Krusell, Mukoyama, Rogerson, and Şahin (2017) and Cairó, Fujita, and Morales-Jimenez (2020) present such richer models of worker flows for all three margins (but do not study congestion dynamics).

J Deriving the Iso-congestion Curve

We generalize the production function in our baseline model and assume a function that takes the following form:

$$Y = (1 - x) \left(\sum_k \alpha_k^c (n_k^c)^\sigma \right)^{1/\sigma} + x \left(\sum_k \alpha_k^{nc} n_k \right).$$

In words, we assume that a share $1 - x$ of workers are subject to short-run congestion and the remaining share x of workers are not subject to congestion in final good production. Alternatively, fraction x of workers enter the k step in a way that replicates the skill structure at the point of hiring. Or, two final goods are produced, which are perfect substitutes but one uses linear production. Search is random, so a given hire is expected to be placed into the two functions with probabilities $1 - x$ and x , respectively.

Marginal Product of Labor. This new production function implies that the expected marginal product of a hire will be, when the congestion hire reaches type- k :

$$p_k = \frac{\partial Y}{\partial n_k} = (1 - x) \underbrace{\alpha_k^c n_k^{\sigma-1} \left(\sum_k \alpha_k^c n_k^\sigma \right)^{1/\sigma-1}}_{=p_k^c} + x \underbrace{\alpha_k^{nc}}_{=p_k^{nc}}.$$

Measure of Congestion. We are interested in how fast the marginal product of labor-type k changes with respect to the mass of employed workers of that particular type. To this end, we use the elasticity of the marginal product of labor with respect to the mass of workers of type k , ε_{p_k, n_k} .

First, we observe that the elasticity of p_k^{nc} with respect to n_k is zero, $\varepsilon_{p_k^{nc}, n_k} = 0$. Second, we calculate the elasticity of p_k^c with respect to n_k

$$\begin{aligned} p_k^c &= \alpha_k^c n_k^{\sigma-1} \left(\sum_k \alpha_k^c n_k^\sigma \right)^{1/\sigma-1} \\ \Rightarrow \varepsilon_{p_k^c, n_k} &= \frac{\partial p_k^c}{\partial n_k} \frac{n_k}{p_k^c} \\ &= (\sigma - 1) \left(1 - \frac{\alpha_k^c n_k^\sigma}{\sum_k \alpha_k^c n_k^\sigma} \right). \end{aligned}$$

Third, we use the property that if $z = x + y$, then the following identity holds for the elasticity of z :

$$\varepsilon_z = \frac{x}{x + y} \varepsilon_x + \frac{y}{x + y} \varepsilon_y.$$

Fourth, using this identity and the fact that $\varepsilon_{p_k^{nc}, n_k} = 0$, we derive our desired elasticity of marginal

product with respect to worker mass:

$$\varepsilon_{p_k, n_k} = (\sigma - 1) \left(1 - \frac{\alpha_k^c n_k^\sigma}{\sum_k \alpha_k^c n_k^\sigma} \right) \frac{(1-x) \alpha_k^c n_k^{\sigma-1} \left(\sum_k \alpha_k^c n_k^\sigma \right)^{1/\sigma-1}}{(1-x) \alpha_k^c n_k^{\sigma-1} \left(\sum_k \alpha_k^c n_k^\sigma \right)^{1/\sigma-1} + x \alpha_k^{nc}}$$

The Iso-congestion Curve. Our calibration ensures that $p_k^c = p_k^{nc} = 1$ for all k , therefore the last term above simplifies to the share of no-congestion workers $1 - x$. Our congestion measure then becomes

$$\varepsilon_{p_k, n_k} = (1-x)(\sigma - 1) \left(1 - \frac{\alpha_k^c n_k^\sigma}{\sum_k \alpha_k^c n_k^\sigma} \right). \quad (\text{A21})$$

Further, as $p_k^c = \alpha_k^c n_k^{\sigma-1} \left(\sum_k \alpha_k^c n_k^\sigma \right)^{1/\sigma-1} = 1$ for all k , we have $\alpha_k^c n_k^{\sigma-1} = \alpha_l^c n_l^{\sigma-1}$. This implies that $\alpha_k^c n_k^\sigma = \alpha_l^c n_l^{\sigma-1} n_k$. Summing over k , we get $\sum_k \alpha_k^c n_k^\sigma = \alpha_l^c n_l^\sigma N / n_l$. Then we obtain $s_l = \frac{n_l}{N} = \frac{\alpha_l^c n_l^\sigma}{\sum_k \alpha_k^c n_k^\sigma}$. Using this result in the elasticity expression above, we finally arrive at

$$\varepsilon_{p_k, n_k} = (1-x)(\sigma - 1)(1 - s_k). \quad (\text{A22})$$

To trace out the iso-congestion curve for $k = 1$, we solve for σ as a function of x given a level of elasticity $\bar{\varepsilon}_{p_1, n_1}$

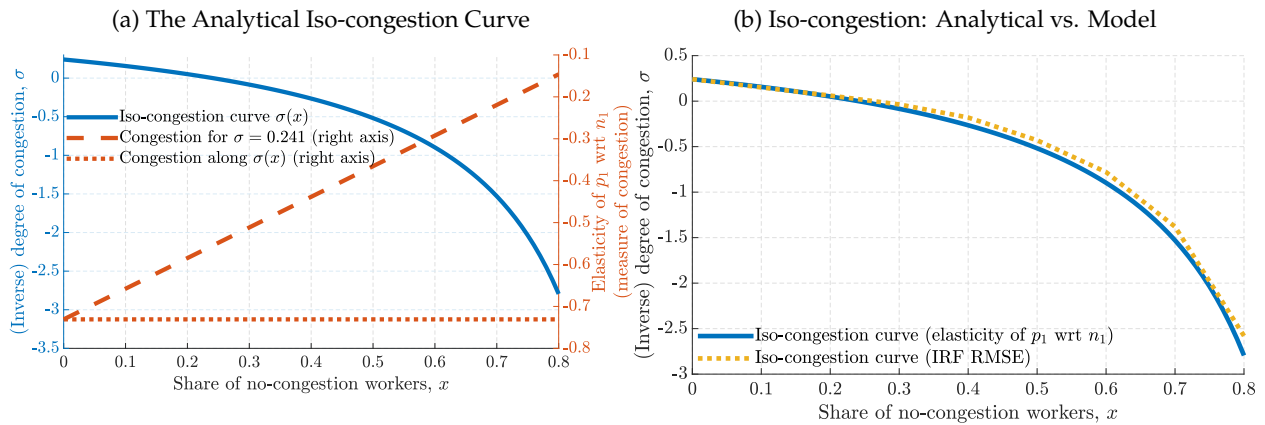
$$\sigma(x) = 1 + \frac{\bar{\varepsilon}_{p_1, n_1}}{(1-x)(1 - s_1)}. \quad (\text{A23})$$

The employment distribution over worker types is characterized by the job finding and separation rates, and the associated laws of motion for employment. Given our calibration strategy (i.e., ensuring $p_k = 1$ for all k), employment share of $k = 1$ workers, s_1 , then stays constant for different levels of the congestion parameter σ .

Figure A19 Panel (a) plots the iso-congestion curve derived in Equation (A23) starting from our baseline calibration of $x = 0$ and $\sigma = 0.241$. The figure makes clear that, as there is more weight on no-congestion workers in final good production, σ needs to be adjusted downward to maintain the same level of congestion as in our baseline calibration. In fact, if $\sigma = 0.241$ is held constant, higher levels of x lead to smaller congestion in production.

Panel (b) superimposes the iso-congestion curve we present in the main text based on the solution to the full dynamic model and on matching the IRF of labor market tightness to the separation rate shock in Figure ?? Panel (a). The figure reveals that, strikingly, the iso-congestion curve we derive analytically overlaps with the one implied by our calibrated model almost perfectly.

Figure A19: Iso-congestion Curves



Notes: Panel (a) plots the analytical iso-congestion curve as a function the share of no-congestion workers in production, x . It also includes the level of congestion as a function of x , as well as the constant level maintained along the iso-congestion curve. Panel (b) compares the analytical iso-congestion curve to the one we obtain solving our dynamic congestion model by matching the IRF of labor market tightness to the separation rate shock in Figure ?? Panel (a).

K Alternative Mechanism: Convex Hiring Costs

Our baseline congestion model obtains congestion in hiring through diminishing returns in the production function. An alternative mechanism of congestion works through a countercyclical *hiring cost* besides the standard DMP vacancy maintenance costs, where, for our purposes, the cost is increasing in UE flows rather than in total hiring:⁹

$$c(UE_t) = c_1 \cdot \left[\left(\frac{UE_t}{UE_{ss}} \right)^{c_2} - 1 \right]. \quad (\text{A24})$$

This cost is zero in steady state; outside of steady state, hiring costs increase in UE flows ($c_1, c_2 > 0$).

The only difference from the standard DMP model is in the free-entry, zero-profit condition, which becomes

$$\frac{\kappa}{q_t} + c(UE_{t+1}) = E_t [\beta(1 - \delta_{t+1})J_{t+1}]. \quad (\text{A25})$$

In turn, we remove worker heterogeneity (essentially setting $\sigma = 1$ and setting the α_k 's to one to yield homogeneous marginal products). Hence, the hiring cost is the only source of congestion, and parameter c_2 guides its degree. We normalize $c_1 = 1$.

The model provides a promising avenue for generating countercyclical congestion by raising the costs of hiring during recessions, when UE flows are high.

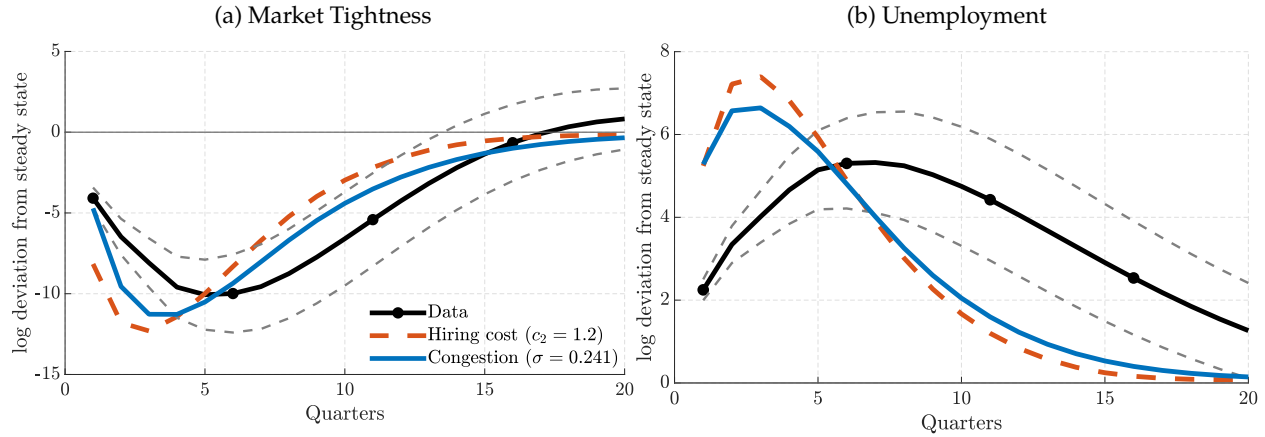
As with the production-function based congestion parameter σ , we now set c_2 such that the model minimizes the RMSE of the response of labor market tightness to separation shocks. Figure A20 shows that the fit of this model is excellent too, closely mirroring the IRF of our main specification in Figure ???. The estimated level of c_2 is 1.2.

The results are presented in Table A11. The model with convex hiring costs can indeed replicate well the volatility of labor market variables. The model also features a robustly negative Beveridge curve and countercyclical UE flows.

Moreover, the model based on convex hiring costs—as our production-based congestion model—is also reasonably sensitive to changes in labor market policies. The elasticity of unemployment with respect to changes in unemployment benefits is 2.59 as is our baseline, production-based congestion model, as it does not rely on a low fundamental surplus to explain labor market volatility. We note that, naturally, the model with convex hiring costs would not generate cyclical displacement costs that are persistent, for lack of cohort effects.

⁹Pissarides (2009); Silva and Toledo (2013) add a fixed costs of hiring, but it is not increasing in the amount of hires.

Figure A20: Impulse Responses to a Separation Shock: Convex Hiring Cost Model



Notes: The figure plots the impulse response functions of market tightness and unemployment to a unit standard deviation separation shock in the data, and the models of congestion through the production function and the convex hiring cost.

Table A11: Business Cycle Properties: Convex Hiring Cost Model

	ALP	f	δ	u	v	θ	UE/E
Standard deviation	0.010	0.061	0.118	0.129	0.096	0.219	0.067
Autocorrelation	0.691	0.855	0.536	0.845	0.856	0.855	0.840
Correlation matrix							
ALP	1						
f	0.505	1					
δ	-0.410	-0.726	1				
u	-0.474	-0.984	0.748	1			
v	0.518	0.967	-0.656	-0.907	1		
θ	0.505	1.000	-0.726	-0.984	0.967	1	
UE/E	-0.346	-0.873	0.316	0.858	-0.846	-0.873	1

Notes: ALP , f , δ , u , θ and UE/E indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and share of new hires in employment, for the model with convex hiring costs. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1,600.

L Historical Decomposition: Additional Material

The main text shows how congestion-only unemployment contributed to the evolution of overall unemployment. In this Appendix, we provide the same exercise also for TFP- and separation-driven unemployment. The estimated time paths of key labor market variables are presented in Figure A21.

The spirit of the decomposition exercise is exactly the same as in the main text and we specify the method below. In particular, we construct counterfactual unemployment rates generated by TFP shocks only, u^z , which would arise in the TFP-shock-only models such as in Shimer (2005); Hall (2005b); Hagedorn and Manovskii (2008) and generated by separation shocks only, u^δ . The corresponding equations that characterize these counterfactuals are, for u^z ,

$$\begin{aligned} u_{t+1}^z &= (1 - f(\theta_t^z))u_t^z + \bar{\delta}(1 - u_t^z), \quad \kappa = q(\theta_t^z)\beta\mathbb{E}_t(1 - \bar{\delta})S_{1,t}^z \\ S_{k,t}^z &= z_t - b + \beta\mathbb{E}_t(1 - \bar{\delta})S_{k+1,t+1}^z - \beta\mathbb{E}_t(1 - \bar{\delta})f(\theta_t^z)\phi S_{1,t+1}^z \text{ for all } k, \end{aligned} \quad (\text{A26})$$

and, respectively, for u^δ ,

$$\begin{aligned} u_{t+1}^\delta &= (1 - f(\theta_t^\delta))u_t^\delta + \delta_{t+1}(1 - u_t^\delta), \quad \kappa = q(\theta_t^\delta)\beta\mathbb{E}_t(1 - \delta_{t+1})S_{1,t}^\delta \\ S_{k,t}^\delta &= \bar{z} - b + \beta\mathbb{E}_t(1 - \delta_{t+1})S_{k+1,t+1}^\delta - \beta\mathbb{E}_t(1 - \delta_{t+1})f(\theta_t^\delta)\phi S_{1,t+1}^\delta \text{ for all } k. \end{aligned} \quad (\text{A27})$$

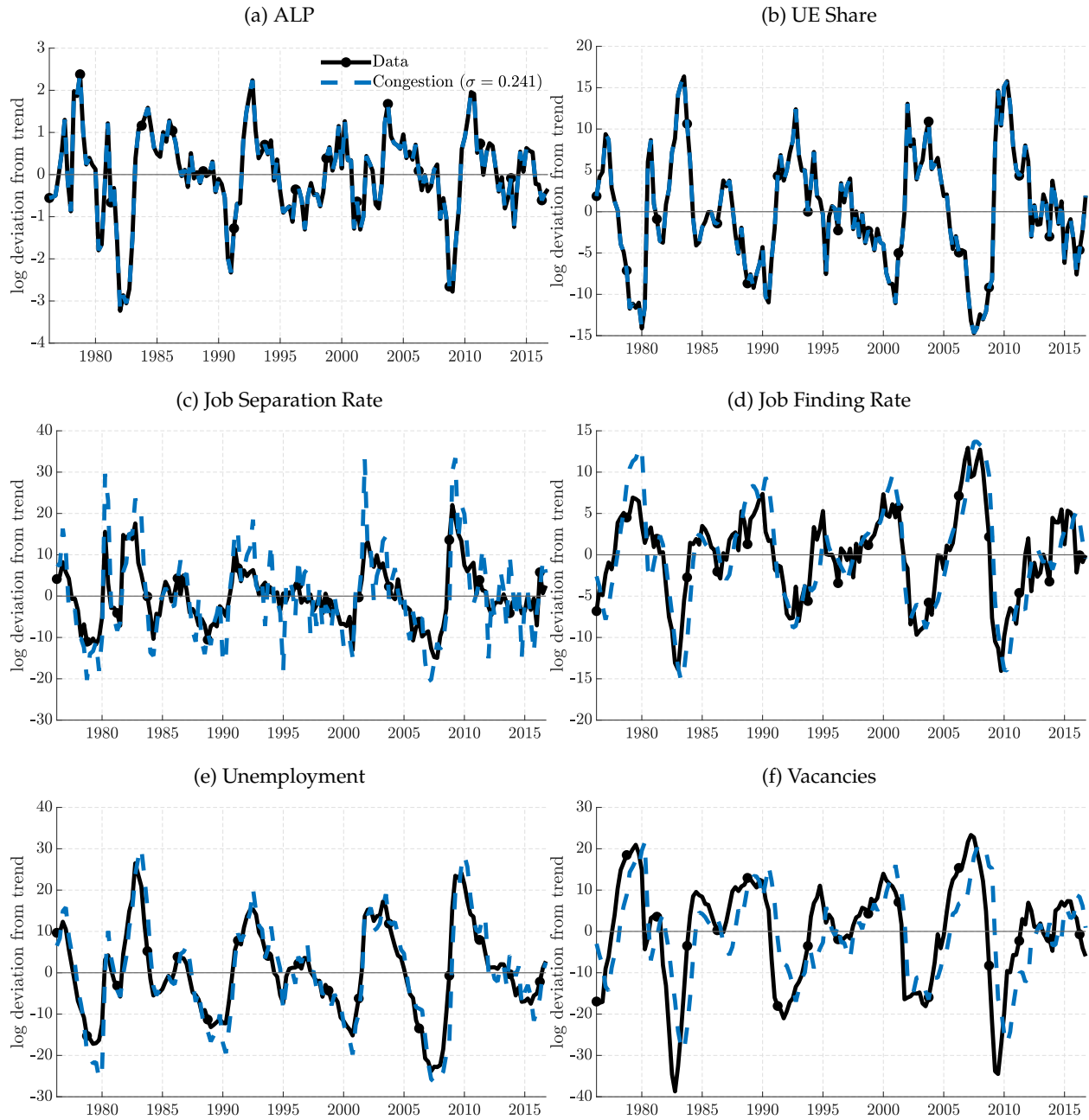
Figure A22 plots the associated time series of these counterfactual unemployment rates together with actual unemployment. Table A12 provides a set of business cycle statistics related to overall unemployment and the three counterfactuals.

Volatility. Table A12 quantifies the role of congestion-driven unemployment in US business cycles, reporting summary statistics of the actual and congestion-only unemployment rates. The congestion-only time series accounts for approximately 30% of the historical unemployment rate fluctuations in the United States. Its standard deviation is around 40% of the empirical one.¹⁰

Persistence and Internal Propagation. Congestion-driven unemployment is considerably more persistent than both TFP- and separation-driven unemployment. Its autocorrelation is 0.950, compared to 0.865 for TFP-driven and 0.825 for separation-driven unemployment rates. This additional persistence arises from the internal propagation mechanisms laid out in Section 5.3.

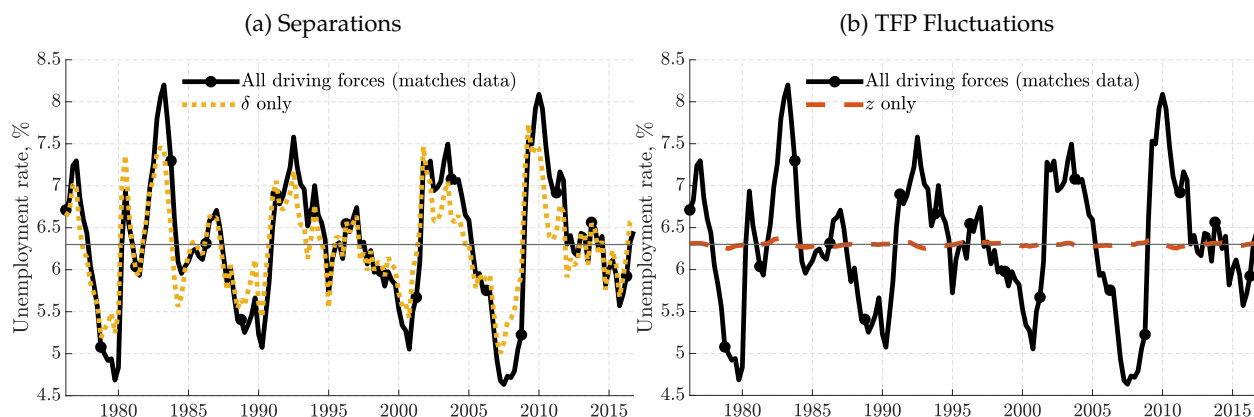
¹⁰As discussed in Section 5.1, our model matches UE flows by estimating a somewhat more volatile separation rate process. In Table A12, this property leads to the model exaggerating the share of unemployment fluctuations due to separation shocks. See Fujita and Ramey (2009) and Shimer (2012) for the empirical contributions of the two transition rates to unemployment fluctuations in the US. A more realistic separation rate process will likely reduce the performance of the model in explaining overall unemployment fluctuations while leaving the contribution of congestion, which manifest itself on the hiring margin, unaffected, as long as that model generates realistic fluctuations in UE flows.

Figure A21: Time Paths of Labor Market Variables



Notes: The figure plots the estimated time paths of labor market variables using the Kalman Filter. Time series are logged and HP-filtered using a smoothing parameter of 1,600.

Figure A22: Unemployment Components



Notes: The figure plots actual, and counterfactual unemployment rates u^z and u^δ estimated using data on the cyclical components of average labor productivity and new hires as a share of employment. The counterfactual unemployment time series are based on Equations (A26) and (A27).

Table A12: Historical Decomposition of Unemployment: Model and Counterfactuals

	Baseline	Congestion only	z only	δ only
Standard deviation	0.124	0.050	0.004	0.088
Contribution to total	1	0.297	0.008	0.657
AR(1)	0.905	0.950	0.865	0.825
$\text{corr}(x, y)$				
Actual	1			
Congestion only	0.729	1		
z only	0.274	-0.264	1	
δ only	0.920	0.411	0.464	1

Notes: This table reports summary statistics for the unemployment rate time series generated using our model (which closely tracks the actual unemployment rate), and the counterfactuals from TFP shocks only, separation shocks only, and congestion only. "Contribution to total" shows $\text{cov}(u_{\text{base.}}, u_{\text{cf.}})/\text{var}(u_{\text{base.}})$, where $u_{\text{base.}}$ is unemployment in our baseline model, while $u_{\text{cf.}}$ is the respective counterfactual unemployment rate.