A Congestion Theory of Unemployment Fluctuations∗

Yusuf Mercan  Benjamin Schoefer  Petr Sedláček
U Melbourne  UC Berkeley  U Oxford & CEPR

January 28, 2021

Abstract

In recessions, unemployment increases despite the—perhaps counterintuitive—fact that the number of unemployed workers finding jobs expands. On net, unemployment rises only because even more workers lose their jobs. We propose a theory of unemployment fluctuations resting on this countercyclicality of gross flows from unemployment into employment. In recessions, the abundance of new hires “congests” the jobs the unemployed fill. The abundance of newly unemployed in the workforce diminishes their marginal product and discourages further job creation. Countercyclical congestion alone explains about 30–40 percent of US unemployment fluctuations. Besides generating realistic labor market volatility, it also provides a unified explanation for the excess procyclicality of new hires’ wages, the cyclical labor wedge, the large earnings losses from job displacement and from graduating during recessions, and the insensitivity of unemployment to labor market policies, such as unemployment insurance.

∗We thank Chris Edmond, Michael Elsby, Domenico Ferraro, Shigeru Fujita, Bart Hobijn, Philip Jung, Fatih Karahan, Simon Mongey, Emi Nakamura, Ayşegül Şahin, Moritz Schularick, Robert Shimer, and Jón Steinsson for useful comments. We also thank seminar audiences at Aarhus University, Arizona State University, Boston University, Erasmus Universiteit Rotterdam, the Federal Reserve Bank of Philadelphia, New York University, UC Berkeley, the University of Melbourne, the University of Adelaide and Waseda University. Sedláček is grateful to the German Research Foundation (DFG), which financially supported early stages of this project 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: petr.sedlacek@economics.ox.ac.uk.
1 Introduction

Recessions are times when labor demand plummets and unemployment increases. Rationalizing why firms are so unwilling to hire away the sudden increase in unemployment remains an actively debated challenge in macroeconomics.\(^1\) We propose a theory of unemployment fluctuations which puts to use a robust, yet somewhat overlooked, empirical fact: in recessions, the number of unemployed who find jobs increases. On net, unemployment rises only because an even larger number of workers lose their jobs. Therefore, recessions are times when newly hired workers from unemployment are abundant in the workforce. In our framework, their abundance in the workforce “congests” the jobs the unemployed fill, diminishing their marginal product and discouraging further job creation. Countercyclical congestion paints a new picture of recessions: rather than asking why firms hire so little, our theory posits that firms have already absorbed so many of the unemployed that the jobs they would fill are already crowded.

We show that countercyclical congestion alone accounts for around 30–40 percent of US unemployment fluctuations and much of its persistence. In addition, our theory provides a unified explanation for a range of other business cycle patterns linked to unemployment: the excess procyclicality of wages of newly hired workers compared to average wages, the countercyclical labor wedge, countercyclical earnings losses from displacement and from labor market entry, and the relative insensitivity of labor markets to policies such as unemployment insurance.

We start our analysis by highlighting that in fact more unemployed find jobs in recessions, despite a drop in the individual probability of finding a job. For instance, during the trough of the Great Recession in 2009, the average number of unemployed workers finding jobs was 20 percent higher compared to the boom year of 2005. We show analytically that the key to understanding these countercyclical unemployment-to-employment (UE) flows is the presence of countercyclical job separations—i.e. the fact that even more people lose their jobs during downturns. Yet, while countercyclical unemployment to employment flows are a robust 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), existing business cycle research has not linked them with firms’ hiring decisions. In fact, frequently used standard search models that assume constant separation rates imply counterfactually procyclical UE flows.

Next, we document that the economy has a limited capacity to absorb new hires—it exhibits congestion in hiring. In particular, we provide new time series evidence showing that firms do not create new jobs in response to increases in unemployment that leave other fundamentals (e.g., productivity) unaffected. Specifically, in response to separation shocks that by construction do not impact average labor productivity on impact, labor market tightness (the ratio of vacancies and unemployment) falls persistently and significantly. This time series fact is robust to accounting for

---

\(^1\)See, 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).
other contemporaneous business cycle shocks. This finding in the aggregate time series is also in line with cross-sectional evidence at the firm level (see, e.g., Doran, Gelber, and Isen, 2020) and local-labor market level (see, e.g., Mian and Sufi, 2014; Gathmann, Helm, and Schönberg, 2018; Mercan and Schoefer, 2020). The negative effect of separations on labor market tightness and job finding stands in sharp contrast to the standard search models, which exhibit no congestion, making firms quickly hire away such an increase in job losers.

Our congestion theory of unemployment integrates both facts into an otherwise standard Diamond-Mortensen-Pissarides (DMP) search and matching model of the labor market. First, shocks to separations generate countercyclical UE flows. Second, we introduce congestion in hiring through an aggregate production function featuring diminishing returns in the size of a given cohort of new hires. The resulting countercyclical congestion rationalizes why in a recession firms do not hire away the additional job losers. We also empirically support this specific congestion mechanism with a range of labor market predictions which, up until now, have not been studied simultaneously.

We formalize diminishing returns in hiring by introducing a constant-returns production function in which different cohorts of hires are imperfect substitutes for one another. For example, different hiring cohorts may be on different rungs of their employers’ internal job ladder, have different experience or firm-specific skill levels, and hence perform different tasks. The key parameter guiding the degree of congestion, and hence the quantitative performance of our model, is the elasticity of substitution between cohorts. With perfect substitution, our framework exactly nests the standard search model (see, e.g., Shimer, 2005). We discipline this congestion parameter by having our model match the empirical impulse response of labor market tightness to a separation rate shock. In the standard, no-congestion DMP model, this response is counterfactually flat (vacancies scale one to one with unemployment), so this impulse response provides a clear target pinning down congestion.

An alternative calibration strategy for our productivity-based congestion mechanism could utilize the excess wage fluctuations of new hires relative to incumbent workers. In the US, new hires’ wages are between two and three times as procyclical as average wages (Pissarides, 2009). Reassuringly, our model predictions fall precisely into this range.

Quantitatively, our model implies that a 10% increase in hires out of unemployment leads to a decline in the cohort’s productivity of 7.6%, corresponding to an elasticity of substitution of worker cohorts of about 1.3. Over the US business cycles, we find that the amplitude of new hires’ productivity remains tightly within an interval of plus and minus 10%, while average labor productivity remains as smooth as in the data.

We find that countercyclical congestion alone accounts for more than 30–40 percent of observed

---

2Specifically, we study 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).

3Here, our calibration strategy echoes the important work by Coles and Moghaddasi Kelishomi (2018) (Table 4), who then propose a model relaxing the free-entry condition.
US unemployment fluctuations, and much of its persistence. The model, which features standard total factor productivity and separation shocks, replicates essentially all the business cycle patterns of labor market variables, unlike the standard DMP model (Shimer, 2005). For example, the standard deviation of labor market tightness is 90 percent of that in the data, and the correlation between unemployment and vacancies, i.e. the Beveridge curve, is $-0.716$ in our model compared to $-0.934$ in the data.

The quantitative success of our model rests on two key features. First, the 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 quantitative performance of our model does not rest on the presence of wage rigidity (see, e.g., Shimer, 2004; Hall, 2005b; Michaillat, 2012; Schoefer, 2015) or a small fundamental surplus (Hagedorn and Manovskii, 2008; Ljungqvist and Sargent, 2017). To address this specific concern, we parameterize our model closely following the choices in Shimer (2005) which, absent countercyclical congestion, would destine the model time series to be counterfactually smooth.

Importantly, we show that even structurally very different congestion mechanisms generate the same amplification as our model as long as they are calibrated to match the empirical degree of congestion (the decline in labor market tightness to separation shocks). We illustrate robustness 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.4

Finally, by offering a new perspective on unemployment fluctuations, our framework offers solutions to three related long-standing macroeconomic challenges. These results provide further external validity for our productivity-based modeling of congestion.

First, countercyclical congestion provides a quantitative explanation of the countercyclical labor wedge, i.e., the gap between the marginal rate of substitution (MRS) between consumption and leisure, and the marginal product of labor (MPL) that is implied by viewing the data through a standard Real Business Cycle model (Chari, Kehoe, and McGrattan, 2007; Shimer, 2009; Karabarbounis, 2014; Bils, Klenow, and Malin, 2018). In fact, the amplification of new hires’ productivity fluctuations in our model is a procyclical multiplier on the standard MPL measure manifesting itself precisely as a countercyclical labor wedge.

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 obtains amplification through more volatile allocative productivity, rather

4Recalls (Fujita and Moscarini, 2017) are an example of hires from unemployment which may not lead to congestion. Moreover, this generalized model structure is reminiscent of models of turbulence without congestion, e.g., [ljungqvist1998european,ljungqvist2004european,den2005turbulence].
than a high elasticity to productivity changes. Hence, we overcome the critique raised by Costain and Reiter (2008), that standard DMP models cannot simultaneously exhibit realistic, productivity-driven, business cycle fluctuations and a low sensitivity to unemployment insurance (UI) generosity.

Our paper relates to Dupraz, Nakamura, and Steinsson (2019); Hall and Kudlyak (2020b,a), who study the dynamics of the gradual reduction in unemployment during recoveries. Our congestion model curbs UE flows and generates persistence in unemployment. A notion of congestion is present in important prior work by Coles and Moghaddasi Kelishomi (2018), who relax the free-entry condition and highlight the role of separation shocks. Hall (2005a) and Engbom (2020) provide 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. Michaillat (2012) presents a model with wage rigidity and diminishing returns in total employment, generating rationing due to wage rigidity. The model does not exhibit congestion in hiring (i.e., it would predict essentially no effect of separation rate shocks on labor market tightness), although in rationalizes a rat-race effect in net employment (Landais, Michaillat, and Saez, 2018). Finally, 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).

In Section 2, we present evidence for the countercyclicality of UE flows and the resulting congestion in hiring. Section 3 presents our model featuring countercyclical congestion. We parameterize and investigate the model’s business cycle performance in Section 4. Section 5 studies three further key macroeconomic implications. Section 6 concludes.

2 Empirical Evidence for Countercyclical Congestion

Countercyclical congestion in new jobs arises from the combination of countercyclical employment shares of recently unemployed workers, and firms’ limited capacity to absorb these UE hires. We now provide empirical evidence for both ingredients.

2.1 Countercyclical Unemployment-to-Employment Flows

The Countercyclical Shift of Employment to the Recently Unemployed. Figure 1 Panel (a) presents our main 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. 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 (or, reported separately, nonemployed) during the previous calendar year. We lead this annual time series by a year to align the reference period, also ensuring consistency with the worker flow analysis we conduct subsequently. The panel also includes the log deviation of unemployment rate from its trend to indicate the state of the business cycle.

Panel (b) illustrates this countercyclical by plotting the log deviation in our employment
Figure 1: Countercyclicality of the Employment Share with Unemployment Past Year

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

(b) Cyclicality: Log Deviations from Trend

(c) Okun’s Law

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. 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).

share measure from its trend (using an HP-filter with a smoothing parameter of 100). Both Panels (a) and (b) further show that this fact is not driven by short unemployment experiences, but is robust to only counting unemployment longer than four weeks, and long-term unemployment totaling at least 26 weeks (after which recalls are essentially zero, Katz and Meyer, 1990; Fujita and Moscarini, 2017). Finally, we quantify the countercyclicality as an Okun’s law (see Panel (c) of Figure 1): the elasticity of the new-hire share in employment with respect to the unemployment rate is 0.493.

Origins: Worker Flows. To understand the countercyclical employment share of workers with recent unemployment documented in Figure 1, we turn to the flow approach to the labor market.
(see, e.g., Davis, Faberman, and Haltiwanger, 2006).

We start by documenting that monthly unemployment-to-employment (UE) worker flows are countercyclical in Figure 2 Panel (a). 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, sample selection and measurement into Appendix A.

UE worker flows expand dramatically during all US recessions since 1976, moving tightly with the unemployment rate. Panel (b) of Figure 2 quantifies this relationship in the form of a scatter plot along with a fitted linear regression line. Expressed as an Okun’s law, the elasticity of UE flows with respect to the unemployment rate is 0.345. That is, UE flows increase by around 3.5 percent when unemployment increases by 10 percent (i.e., around 0.63 ppt from an average baseline 6.3 percent unemployment rate 1976–2019). Appendix Figure A3 Panel (a) reports this elasticity for UE hires as a share of employment, which implies an elasticity of 0.432, consistent with the result depicted in Figure 1 Panel (c). Appendix B shows that countercyclical UE are a feature across the OECD. The countercyclicality of UE flows has been documented as a stylized fact (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).
The Role of Countercyclical Separations  Next, to shed light on the proximate causes behind the countercyclical employment share of UE hires, we decompose UE flows into contributions from two worker transition rates in a two-state labor market model featuring employment and unemployment, abstracting from labor force participation. Each period, a fraction $\delta$ ("separation rate") of employed workers separate into unemployment, and a fraction $f$ ("job finding rate") of unemployed job searchers find, and accept, a job.\(^5\) This bathtub model of "ins" and "outs" of unemployment implies a steady-state unemployment rate given by

$$u = \frac{\delta}{\delta + f}. \quad (1)$$

UE flows per period are given by the number of job seekers $U$ times the individual job finding rate $f$,

$$UE = f \cdot U. \quad (2)$$

Hence, the percent change in UE flows, by totally differentiating Equation (2), is equal to

$$\frac{dUE}{UE} = \frac{df}{f} + \frac{dU}{U}. \quad (3)$$

Equation (3) shows that for UE flows to increase together with unemployment, unemployment must increase disproportionately more than the job finding rate falls in a recession.

Using the expressions above and normalizing the (constant) labor force to 1 (such that $u = U$), we can recover the elasticity of UE flows with respect to the unemployment rate depicted in Figure 2 Panel (b) as follows:

$$\frac{dUE/UE}{du/u} = \frac{df/f}{du/u} + 1 = \frac{1}{(1 - u)\left[-1 + \frac{d\delta/\delta}{df/f}\right]} + 1, \quad (4)$$

where we use the fact that $\frac{du}{u} = (1 - u) \left[-\frac{df}{f} + \frac{d\delta}{\delta}\right]$ implied by Equation (1). Equation (4) reveals that the sign of the UE elasticity is a priori ambiguous. If separations were constant—as is a common assumption in search models (see a discussion in e.g., Shimer, 2005)—then UE flows are procyclical, namely $\frac{dUE/UE}{du/u} = -\frac{u}{1-u}$. However, if separations are time varying and sufficiently countercyclical (i.e., if $\frac{d\delta/\delta}{df/f} < -\frac{u}{1-u}$), UE flows turn countercyclical.

In the US, separations are indeed sufficiently countercyclical to generate countercyclical UE flows. In Figure 3 Panel (a), we plot the detrended time series of both the job finding and job separation rates. Their correlation is strongly negative at $-0.717$. Both time series are also relatively volatile, with standard deviations of 0.070 and 0.068, respectively. These values imply that $\frac{d\delta/\delta}{df/f} \approx -0.698$, which is considerably below the threshold $-u/(1-u) \approx -0.067$.

\(^5\)In the data, and later on in the model, we specify discrete-time transition probabilities while using the conventional term “rates” interchangeably.
Figure 3: Transition Rates and Counterfactual Worker Flows

Notes: Panel (a) plots log deviations of quarterly-averaged monthly UE and EU rates from their trends. Panel (b) 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. 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.

We illustrate the importance of separations in generating countercyclical UE flows in Panel (b) of Figure 3. We obtain a counterfactual UE flow time series based on the law of motion for unemployment, using the observed job finding rate yet holding the separation rate at its sample average $\bar{\delta}$. 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). Intuitively, the reason why separations drive UE dynamics can be seen from combining Equation (3) with the total derivative of Equation (1), which yields $\frac{d\bar{UE}}{dU} = u \frac{df}{dU} + (1 - u) \frac{d\bar{\delta}}{dU}$. Here, in percent terms, movements in the separation rate $\delta$ contribute to UE flows by more than $\frac{1 - u}{u} \approx 15$ times the amount the job finding rate $f$ does.

Time Aggregation Adjustment. For consistency with the discrete time model that we present below, the empirical transition rates are not adjusted for time aggregation. 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 C, we find very similar results for the cyclical behavior of these UE flows adjusted for such time aggregation.

---

6Specifically, we iterate on the law of motion for unemployment given by $\bar{U}_{t+1} = (1 - f_t)\bar{U}_t + \bar{\delta}(L_t - \bar{U}_t)$ to construct the counterfactual unemployment time series $\bar{U}_t$ over our sample, where $f_t$ and $L_t$ denote the observed job finding rate and labor force in month $t$. Then our counterfactual time series for UE flows is $\bar{UE}_t = f_{t-1}\bar{U}_{t-1}$. 

8
**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). Total (rather than those only out of unemployment) hires are not countercyclical. We view hires from unemployment as filling jobs which are fundamentally different from those filled by job-to-job movers. Therefore, we focus on countercyclical congestion in jobs filled by workers hired out of unemployment, their share in employment, and (their effect on) flows between unemployment and employment.

**Unemployment vs. Nonemployment.** In Appendix Figures A1 to A4, 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 in order to match our empirical calibration targets. In a model extension, we will also 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 D, we show that our results are robust to the alternative smoothing parameter. Most importantly, the elasticity of UE flows with respect to the unemployment rate is identical (0.348 vs 0.345) as is the employment share of new hires out of unemployment (0.433 vs 0.432).

### 2.2 Evidence for Congestion Effects in Hiring

We now provide evidence for congestion effects, i.e., the limited capacity of the economy to absorb new hires compared to a no-congestion benchmark. We begin by defining congestion in hiring, showing how the dramatic differences in labor market dynamics with and without congestion. Next, we review quasi-experimental evidence for congestion in hiring at the firm- and local-labor market level. Finally, we end this subsection providing novel time-series evidence for congestion at the aggregate level.

**Defining Congestion.** We define our congestion concept as the economy’s limited capacity to absorb—by means of UE flows—“pure disturbances” in the unemployment pool (i.e., that leave fundamentals such as productivity and the discount factor constant). The standard DMP model features no congestion in that sense whatsoever. To fix ideas, we now juxtapose the dynamics in this standard, no-congestion model with an extreme, full-congestion benchmark. In the full-congestion benchmark, the economy cannot respond at all to short-run spikes in unemployment, and UE flows remain fixed.
In both models, labor market tightness \( \theta = \frac{v}{u} \), the ratio of vacancies \( v \) to the unemployed \( u \), determines the job finding rate \( f(\theta) \). Having in mind a constant-returns-to-scale matching function, we assume that \( f'(\theta) > 0 \) and \( f''(\theta) < 0 \). Finally, in both economies unemployment evolves according to the following law of motion:

\[
\begin{align*}
    u_{t+1} &= (1 - f(\theta_t)) u_t + \delta_{t+1}(1 - u_t) .
\end{align*}
\]  

**Standard Labor Market Adjustment Without Congestion.** In the labor market without congestion, of which the standard DMP economy is an example, hiring (vacancy posting) is determined by a labor demand condition in which equilibrium vacancies respond only to changes in the benefits or costs of hiring. In response to pure shifts in unemployment, equilibrium vacancies simply scale one to one, such that their ratio with unemployment (labor market tightness \( \theta \)) remains fixed. This property, in turn, implies that also the job finding rate \( f(\theta) \) remains fixed, enabling the economy to quickly absorb the (pure) spike in unemployment through a spike in hiring. The adjustment paths for this economy can be conveniently summarized analytically, assuming a perfectly transitory, positive separation shock hitting in \( t = 0 \), i.e., \( \delta_0 = \delta_{ss} + \delta \delta_0 \) and \( \delta_t = \delta_{ss} \forall t > 0 \):\(^7\)

\[
\begin{align*}
    \theta_t &= \theta_{ss} , \quad v_t = \theta_{ss} \cdot u_t , \quad UE_t = u_t \cdot f(\theta_{ss}) \\
    u_{t+1} &= u_t - f(\theta_{ss})u_t + (1 - \delta_{ss})(1 - u_t) .
\end{align*}
\]  

We illustrate these no-congestion labor market dynamics in the red dashed lines in Figure 4, plotting the theoretical impulse responses to an increase in the unemployment pool brought about by a one-time, perfectly transitory increase in the separation rate \( d \delta_0 = \delta_0 - \delta_{ss} \). Upon impact, unemployment incipiently increases by \( d \delta_0 \cdot u_{ss} \), the inflow from the extra job losers. Immediately, however, vacancies exhibit a tantamount upward spike—so vacancies and unemployment move into the same direction. The vacancy surge keeps labor market tightness \( \theta \) and the job finding rate \( f(\theta) \) constant. Higher unemployment combined with a constant job finding rate, in turn, leads UE flows to spike, which is exactly the mechanism that achieves nearly immediate convergence back to steady state—absent congestion in UE hiring.

**Congested Labor Market Adjustment.** A counterexample to the no-congestion model is one in which the economy cannot easily absorb increases in unemployment, including those following a separation shock. In the extreme case of full congestion, UE flows remain constant. We can again analytically solve for the transition path of unemployment in this model. These paths make immediately clear that, in the presence of congestion, labor market tightness and the job finding

\(^7\)Foreshadowing our quantitative exercises, we adopt the standard Cobb-Douglas matching function, \( M(u,v) = \overline{\mu} u^{1-\mu} v^\mu \) (for which \( f(\theta) = \mu \theta^{1-\mu} \), where \( \overline{\mu} \) denotes matching efficiency and \( \mu \) is the matching elasticity). We set \( \overline{\mu} = 0.57 \) and \( \mu = 0.72 \) based on our preferred model calibration.

\(^8\)The half life \( t_{0.5}^{uc} \), i.e., the time it takes to arrive at \( \frac{u_t - \delta_{ss}}{u_{ss} - u_{ss}} = 0.5 \), in the no-congestion model is \( u_t - u_{ss} = (1 - (f_{ss} + \delta_{ss})) \cdot (u_0 - u_{ss}) \Rightarrow \frac{u_t - u_{ss}}{u_{ss} - u_{ss}} = (1 - (f_{ss} + \delta_{ss})) \), and hence \( t_{0.5}^{uc} = \log(0.5)/\log(1 - (f_{ss} + \delta_{ss})) \). Since US labor markets are fluid, with quarterly \( f_{ss} \approx 0.570 \) and \( \delta_{ss} \approx 0.042 \), this half life is short, around 0.731 quarters.
rate must \textit{fall} when unemployment inflows increase:

\[ UE_t = f(\theta_t)u_t = UES, \quad \theta_t = f^{-1}(UES/u_t), \quad v_t = \theta_t \cdot u_t \]
\[ u_{t+1} = u_t - f(\theta_t)u_t + (1 - \delta_{ss})(1 - u_t). \]  \hfill (7)

Figure 4 plots, with the dotted yellow lines, the transition paths of this full congestion case.\(^9\) While upon impact, unemployment increases by the same amount as in the no-congestion benchmark, the transition dynamics differ dramatically. UE flows are constant, rather than increasing sharply. To achieve constant gross hiring (UE flows) in the face of an abundance of unemployed, the job finding rate and hence labor market tightness must \textit{fall}. This drop can only be brought about by a drop in vacancies. Consequently, unemployment recovers extremely slowly.

\[^9\]Following similar steps as in the no-congestion case, we derive the half life of the unemployment recovery in the full-congestion model, \( t_{0.5}^{fc} \), as follows: \( u_t - u_{ss} = (1 - \delta_{ss})t \cdot (u_0 - u_{ss}) \Rightarrow \frac{u_t - u_{ss}}{u_0 - u_{ss}} = (1 - \delta_{ss})^t \), and hence \( t_{0.5}^{fc} = \log(0.5)/\log(1 - \delta_{ss}) \). Calibrated to the US average \( \delta = 0.042 \), this half life is around 16 quarters.
Review of Cross-Sectional, Quasi-Experimental Evidence. Before we present our aggregate time series evidence for congestion, we argue that there exists a considerable amount of compelling firm-level and local-labor market evidence for congestion in hiring.

First, at the firm level, Doran, Gelber, and Isen (2020) draw on quasi-experimental variation in recent hires arising from US visa lotteries. They find that one exogenously assigned new hire (more than) fully crowd out any additional subsequent hiring into that job type—which would imply full congestion at the firm level. Since the hiring response is concentrated in specific (new) job types, this evidence is consistent with a target employment count in a narrowly defined category of entry level jobs, rather than a total employment target. Such hiring targets are also consistent with qualitative evidence on the organization of work that renders entry-level jobs imperfect substitutes for higher-tier jobs (see, e.g., the “ports of entry” described in Doeringer and Piore, 1985). The congestion mechanism adopted in our theoretical framework will reflect such features.

Second, cross-sectional evidence from local labor market adjustment is consistent with congestion in hiring. Studying the degree to which vacancy posting complies with free entry in the context of vacancy chains, Mercan and Schoefer (2020) review 15 studies of local labor market adjustment in response to firm- or industry-specific shocks to local employment. Their meta-analysis documents very limited short-run employment spillovers from firms directly affected by some labor demand shifters in a local labor market onto peer firms not directly impacted by the first group’s labor demand shifters. 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). Therefore, these studies also point to the presence of congestion in hiring with local labor markets having a limited capacity to absorb spikes in unemployment.

Aggregate Time Series Evidence. Our main quantitative evidence for congestion in hiring implements the thought experiment presented in Figure 4 in US time series data. Using a vector autoregression (VAR) model, we study the response of labor market tightness to a separation-shock induced expansion in unemployment. Specifically, we study the behavior of two sets of endogenous variables given by the vector:¹⁰

\[ y_t = [\ln ALP_t, \ln \delta_t, \ln x_t], \]

¹⁰Coles and Moghaddasi Kelishomi (2018) also study the responses of labor markets to separation shocks. We argue that their limitation of free entry also constitutes a congestion mechanism, and resembles our additional mechanism of convex (UE) hiring costs which, however, misses some of our key results pertaining to workers’ wage fluctuations and earnings losses from displacement.
where $ALP$ is average labor productivity (measured as output per worker in the non-farm business sector), $\delta$ is the separation rate (EU flows divided by beginning-of-period employment), and $x$ denotes either labor market tightness (vacancies from Barnichon, 2010, divided by unemployment) or the unemployment rate. To be consistent with our subsequent quantitative framework and due to data limitations ($ALP$ is measured on a quarterly frequency), we convert the monthly job separation rate to a quarterly measure.

We then estimate the following VAR model for each endogenous variable vector $y_t$:

$$y_t = c + A(L)y_{t-1} + v_t,$$

where $c$ is a constant term, $A(L)$ is a lag polynomial, and $v_t \sim (0, \Omega)$ is a vector of error terms with variance-covariance matrix $\Omega$. 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 $\Omega$). Our timing assumptions are that $ALP$ has a contemporaneous effect on both $\delta$ and $x$. In contrast, $\delta$ only has a contemporaneous effect on $x$ and $x$ affects the endogenous variables only with a lag. We then study impulse responses to an orthogonalized shock to $\delta$, to isolate the effect of movements in job separations from that of productivity fluctuations.

Figure 5 plots the empirical impulse response functions of labor market tightness (Panel (a)) and unemployment (Panel (b)) to a separation shock. We also report the two counterfactual benchmark responses from an economy with no congestion (red dashed line), and full congestion in hiring (yellow dotted line).

The data clearly reject the insensitivity of labor market tightness predicted by the no-congestion benchmark. The empirical response is significantly negative and persistent (Panel (a)). That is, vacancies do not quickly and sufficiently expand to absorb the newly unemployed workers. The resulting drop in the job finding rate, paired with the increase in separations, triggers a large and persistent increase in unemployment (Panel (b)). These empirical patterns are absent in the standard DMP, no-congestion model.

Of course, quantitatively, the empirical responses still lie in between the no-congestion and full-congestion extremes. Therefore, in our quantitative model, we pin down the precise degree of congestion by having our model match the empirical market tightness response to a separation shock depicted in Figure 5 Panel (a).

---

11This orthogonality with productivity holds exactly in the first period. In Appendix Figure A10, we present the IRFs of $ALP$ to the $\delta$ 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.

12The counterfactual responses of no- and full-congestion mimic those in Figure 4 but use the estimated IRFs of labor market tightness to the empirical separation shock.
Figure 5: Congestion: Empirical Impulse Responses to a Separation Shock

(a) Labor Market Tightness

(b) Unemployment

Notes: Panel (a) plots the impulse response of labor market tightness to a unit standard deviation job separation shock using a three-variable VAR, identified off a Cholesky decomposition. Panel (b) plots the impulse response of the unemployment rate to a separation shock. The dashed lines are one standard deviation confidence bands. The figure also includes two extreme benchmarks, no- and full-congestion impulse responses to the same shocks.

Discussion of VAR Identification. More so than in the cross-sectional design, 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).

To address these concerns, Appendix E assesses the role of omitted shocks in the 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 other macroeconomic shocks leaves the specific time-path of our separation shocks essentially unchanged. We conclude that the leading candidates of observable shocks are unlikely to confound our estimation of the congestion dynamics.

\[13\] An 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.
3 A Search Model with Countercyclical Congestion

We now integrate countercyclical congestion into an otherwise standard DMP model. In Section 4, we calibrate the model and study its quantitative performance for core labor market fluctuations, while Section 5 shows that our framework provides a unified explanation for a range of other labor market phenomena.

To model congestion in hiring from unemployment, we add two ingredients into the canonical DMP framework. First, we generate countercyclical UE flows, by adding countercyclical separations. Second, to obtain congestion dynamics, our model features diminishing returns in new hires—arising from imperfect substitution between hiring cohorts in an aggregate production function. (We also show robustness to congestion from convex hiring costs.) When UE flows rise, as they do in recessions, new hires become relatively abundant. 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 and their imperfect substitutability in production. This feature generates the diminishing returns in hiring, which acts as the source of congestion in our model.

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

Figure 6 summarizes how worker types evolve in our setting during employment and unemployment spells. Each period a worker is employed, she moves up 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 $k_u(k) \in \{0, 1, \ldots, k - 1\}$ determines the size of the downgrade as a function of current type $k$. This setup nests various possibilities 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 $k_u(k) = k - 1$. Specific processes that could underlie these processes include general or firm-specific human capital accumulation and loss while unemployed (Ljungqvist and Sargent, 1998, 2004; den Haan et al., 2005). However, for tractability and the benefit of a direct comparison to the standard DMP model—which, as will become clear, our framework nests—we abstain from a more detailed modelling of a particular skill process.

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 unite of intermediate goods. We denote the stock of type-$k$ workers (and hence intermediate inputs) by $\{n_k\}_{k=1}^K$. Intermedi-
ate 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 aggregate total factor productivity (TFP) \( z \). The aggregate production function is given by

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

where \( \alpha_k \) is a type-specific productivity shifter associated with type \( k \), and \( \sigma \) 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, \( \frac{1}{1-\sigma} \). The standard DMP model emerges a special case when worker types are perfect substitutes for one another (and no differences in productivity weights \( \alpha \)). This nesting result will permit us to isolate the congestion mechanism.

The competitive price for each intermediate input \( k \) reflects the marginal product of labor-type \( k \) engaged in that good’s production:

\[
p_k = \frac{\alpha_k n_k^{\sigma-1} Y}{\sum_{l=1}^{K} \alpha_l n_l^\sigma} = \frac{\alpha_k s_k^{\sigma-1} 1}{\sum_{l=1}^{K} \alpha_l s_l^\sigma} \frac{Y}{N},
\]

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 (10) 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.** Consider the specific case that upon job loss, workers fully downgrade to \( k = 1 \), i.e., \( k_u(k) = k - 1 \) for all \( k \). In this case, all unemployed workers become the same type. Hence, all UE hires are also the same type, and will climb the

---

\(^{14}\)In Appendix F, we present a generalization that allows for perfect substitution between subsets of worker types, thereby permitting one to generalize the skill accumulation and decumulation processes further.
worker-type ladder as one cohort. This case permits an easy representation of new hires’ marginal product of labor, namely \( p_{k=1} \).

Figure 7 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 \( \alpha_k \) such that steady-state marginal products are normalized to one for all worker types, i.e., \( p_k = 1 \) for all \( k \), for each \( \sigma \) 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. 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. The blue solid line is the productivity of new hires, which falls (rises) sharply when new hires become abundant (scarce). Specifically, an increase in the share of new hires of 10 percent (that is, 0.4ppt off the baseline of 4 percent) lowers productivity by around 7.6 percent (the local slope of \( 1 - 0.241 = 0.759 \)). As another way to judge \( \sigma \), 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.

In the data, this mechanism provides large fluctuation of new-hire compared to average productivity. Panel (b) of Figure 7 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 \( p_{1,t} = \alpha_1 s_{1,t}^{-1} ALP_{t} \frac{1}{\sum_{k=1}^{K} \alpha_k s_{k,t}} \), where \( ALP_t \) is the observed average labor productivity. 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 (\( SD(ALP) = 0.010 \)) used in the existing 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 amplitudes of new-hire productivity remain tightly within an interval of plus and minus 10%.

**Discussion: Segmentation of Cohorts** The tractable assumption that hiring cohorts remain segmented throughout their tenures, even, e.g., 20 years into the job, may appear unappealing.\( ^{15} \)

\( ^{15} \)For 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 hires share while ensuring that the shares of the other types \( k > 1 \) drop accordingly.
Figure 7: Flow Productivity and The Size of the Hiring Cohort

(a) Productivity vs. New-Hire Share

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 $\sigma$. Steady-state average labor productivity and each type’s marginal product are normalized to one for both calibrations of $\sigma$. Panel (b) plots the empirical US time series for average productivity and new-hire productivity. Both time series are in logs and detrended using an HP-filter with a smoothing parameter of 1,600.

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 diminishing returns parameter $\sigma$ to match the empirical congestion response in hiring. Finally, the new-hire productivity time series exhibits persistence, which compresses productivity differentials between adjacent cohorts.

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 they sell in a perfectly competitive market to a final good produce. The latter “retailer” bundles the intermediate goods into a final consumption good using the technology in Equation (9) with total factor productivity (TFP) $z$. 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 is $q(\theta) = \frac{M}{v} = M(1/\theta, 1)$.

**Separations.** Each period, active matches separate with exogenous but time-varying rate $\delta$. These separations are an ad-hoc event rather than arising from endogenous decisions between the worker and firm in response to shocks to surplus. While we do not provide this extension, 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 5).\(^{16}\)

**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 progress of a worker through the original employer’s job ladder. In the broadest sense, one could think of the job ladder 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 (skill) type evolution 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).\(^{17}\)

**Aggregate State Variables.** The economy is subject to aggregate shocks, namely to the job separation rate $\delta$ 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 $\delta$ are realized. Worker-firm matches (both those active last period and those formed last period) are destroyed at rate $\delta$, in which case the worker becomes 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

\(^{16}\)An 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).

\(^{17}\)If the mechanism worked through the job ladder or job types 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.
workers downgrade by $k_u$ 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:

\[
\begin{align*}
    u_{k\rightarrow k_u(k),t} &= \left(1 - f(\theta_{t-1})\right) u_{k,t-1} + \delta_t e_{k\rightarrow k_u(k),t} \quad \text{for all } k \\
    e_{k\rightarrow k_u(k),t} &= (1 - \delta_{t-1}) e_{k\rightarrow k_u(k)-1,t-1} + f(\theta_{t-1}) u_{k,t-1} \quad \text{for all } k,
\end{align*}
\]  

(11)

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 $u_t = \sum_k u_{k,t} = L - (1 - \delta_t) \sum_k e_{k,t}$.

**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.

**Worker and Firm Problems.** We cast the worker and firm value functions 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

\[
W_{k,t} = w_{k,t} + \beta E_t \left[ (1 - \delta_{t+1}) W_{k+1,t+1} + \delta_{t+1} U_{k+1,t+1} \right],
\]  

(12)

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

\[
U_{k,t} = b + \beta E_t \left[ f(\theta_t)(1 - \delta_{t+1}) W_{k-1,k_u(k),t+1} + (1 - f(\theta_t)(1 - \delta_{t+1}) U_{k-1,k_u(k),t+1} \right],
\]  

(13)

where $b$ is the flow value of unemployment.\(^{18}\) If the worker contacts a firm and does not separate at the beginning of the next period, she becomes employed next period. 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.

\(^{18}\)We 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.
Firm problems mirror that of the workers. The value of a vacancy is

\[ V_t = -\kappa + q(\theta_t)\beta E_t \left[ (1 - \delta_{t+1}) \sum_k \frac{u_{k,t}}{u_t} J_{k-k_a(k),t+1} + \delta_{t+1} V_{t+1} \right], \]  

(14)

where the firm pays flow cost \( \kappa \) to maintain the vacancy and \( \sum_k \frac{u_{k,t}}{u_t} J_{k-k_a(k),t+1} \) 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

\[ J_{k,t} = p_{k,t} - w_{k,t} + \beta E_t \left[ (1 - \delta_{t+1}) J_{k+1,t+1} + \delta_{t+1} V_{t+1} \right], \]  

(15)

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

\[ S_{k,t} = W_{k,t} - U_{k,t} + J_{k,t} - V_t. \]  

(16)

The individual value functions in Equations (12)–(15) and the definition of surplus in Equation (16) yield the following surplus value:

\[ S_{k,t} = p_{k,t} - b + \beta E_t \left[ (1 - \delta_{t+1}) S_{k+1,t+1} - f(\theta_t)(1 - \delta_{t+1}) \phi S_{k-k_a(k),t+1} + U_{k+1,t+1} - U_{k-k_a(k),t+1} \right]. \]  

(17)

The value of unemployment can be expressed in terms of match surplus as follows

\[ U_{k,t} = b + \beta E_t \left[ f(\theta_t)(1 - \delta_{t+1}) \phi S_{k-k_a(k),t+1} + U_{k-k_a(k),t+1} \right]. \]  

(18)

The wage for worker type \( k \) is determined period-by-period and according to generalized Nash bargaining:

\[ w_{k,t} = \arg \max_{W_{k,t} - U_{k,t}} \phi (J_{k,t} - V_t)^{1-\phi}, \]  

(19)

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

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

(20)

In words, the worker captures a constant share \( \phi \) of the total match surplus, and the firm captures the rest.
Free entry of firms pins down $V_t = 0$ for all $t$. Equation (14) therefore implies

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

(21)

**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 market tightness $\theta$, such that:

- match surplus $S_k$ solves the Bellman equation in Equation (17) for all $k$,
- unemployment value $U_k$ solves the Bellman equation in Equation (18) for all $k$,
- intermediate goods prices $p_k$ satisfy Equation (10) for all $k$,
- masses of (un)employed, $u_k$ and $e_k$, follow the laws of motion in Equation (11) for all $k$,
- end-of-period intermediate goods are given by $n_k = (1 - \delta)e_k$ for all $k$,
- market tightness $\theta$ solves the free-entry condition in Equation (21),
- exogenous state variables $z$ and $\delta$ follow stochastic processes specified in Section 4.

## 4 Quantitative Analysis: Labor Market Fluctuations with Countercyclical Congestion

We now study the model quantitatively. We first discuss our calibration strategy, and then analyze the business cycle properties of the calibrated model. Section 5 then shows how our model simultaneously provides an explanation for a range of other macroeconomic patterns connected to unemployment fluctuations that have, otherwise, been difficult to rationalize within a single model.

### 4.1 Model Parameterization

Table 1 summarizes the model parameters and the targets we use to discipline them. Appendix G 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). With congestion, we additionally discipline parameters of the aggregate production function—the congestion parameter $\sigma$, and the relative weights of different types in production, $\alpha_k$.

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.\footnote{To be consistent with our discrete time model, transition rates are not adjusted for time aggregation. Appendix C reports how our measured flows compare to time-adjusted flows and that our data is essentially the same as that used by Shimer (2012).}

We set the discount factor to $\beta = 0.99$, which yields an annual real interest rate of about 4 percent. 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, $\phi = \mu$.\footnote{The 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).} Finally, the vacancy posting cost, $\kappa$, 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 percent, 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 more volatile allocative productivity of new hires.

In addition, we ensure that steady-state surpluses are identical across all model variants (e.g., when considering different values of $\sigma$) by setting the type-specific productivity weights $\alpha_k$ such that $p_k = 1$ for all $k$ in steady state. We report details on this procedure in Appendix H.

**Worker Type Evolution: Full Downgrading to $k = 1$ (W.L.O. Quantitative G).** 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 attains this tenure level given the separation rate.

In our baseline specification—without loss of 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 also permit gradual skill decline, although in these models all worker skill types are perfect substitutes in production).

In Section 4.2, 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.
Table 1: Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target/Source</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Congestion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discount factor (β)</td>
<td>0.99</td>
<td>Annual interest rate</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Matching elasticity (μ)</td>
<td>0.72</td>
<td>Shimer (2005)</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Matching efficiency (m)</td>
<td>0.57</td>
<td>Job finding probability</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>Bargaining power (η)</td>
<td>0.72</td>
<td>Hosios condition</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Unemployment flow value (b)</td>
<td>0.39</td>
<td>Avg. replacement rate</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>Vacancy posting cost (κ)</td>
<td>0.21</td>
<td>Normalization θ = 1</td>
<td>–</td>
<td>1.00</td>
</tr>
<tr>
<td>Productivity shock, mean (z)</td>
<td>1</td>
<td>Normalization</td>
<td>–</td>
<td>1.00</td>
</tr>
<tr>
<td>Productivity shock, st. dev. (σz)</td>
<td>0.008</td>
<td>St. dev. of ALP</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>Persistence of ALP (ρz)</td>
<td>0.956</td>
<td>Persistence of ALP</td>
<td>0.74</td>
<td>0.69</td>
</tr>
<tr>
<td>Separation shock, mean (δ)</td>
<td>0.037</td>
<td>Unemployment rate</td>
<td>0.063</td>
<td>0.063</td>
</tr>
<tr>
<td>Separation shock, st. dev. (σδ)</td>
<td>0.107</td>
<td>St. dev. of UE/E</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>Persistence of UE/E (ρδ)</td>
<td>0.709</td>
<td>Persistence of UE/E</td>
<td>0.84</td>
<td>0.74</td>
</tr>
<tr>
<td>Correlation of shocks to z and δ</td>
<td>−0.505</td>
<td>corr(ALP, δ)</td>
<td>−0.41</td>
<td>−0.41</td>
</tr>
<tr>
<td>Elasticity of substitution b/w workers (σ)</td>
<td>0.241</td>
<td>Impulse response of θ to δ, see Figure 9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Relative productivities of worker types (ak)</td>
<td>see Appendix H</td>
<td>pok = 1 for all k in steady state</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Parameter values for both the baseline "congestion" model (σ = 0.241) and for the standard "no-congestion" model (σ = 1). Model-implied 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.
Aggregate Shocks. Aggregate productivity \( z \) and job separation rate \( \delta \) follow AR(1) processes in logs,

\[
\ln(z_{t+1}) = (1 - \rho_z) \ln(\overline{z}) + \rho_z \ln(z_t) + \sigma_z \varepsilon_z^{z}_{t+1} \tag{22}
\]

\[
\ln(\delta_{t+1}) = (1 - \rho_\delta) \ln(\overline{\delta}) + \rho_\delta \ln(\delta_t) + \sigma_\delta \varepsilon_\delta^{\delta}_{t+1}, \tag{23}
\]

where \( \overline{z} \) and \( \overline{\delta} \) 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. While average productivity is normalized to one, the average separation rate \( \overline{\delta} \) is set such that the model matches an average unemployment rate of 6.3 percent for our sample period of 1976–2019. In order to pin down the persistence and volatility parameters, we target the observed autocorrelation and standard deviation of average labor productivity (real output per worker in the non-farm business sector) and the share of UE flows in employment.\(^{21}\) Finally, we let the correlation between \( \varepsilon_z \) and \( \varepsilon_\delta \) 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 \( \sigma \), which we describe below, as the behavior of UE flows is an equilibrium outcome.

Disciplining Congestion Parameter \( \sigma \): Impulse Response of Labor Market Tightness to Separation Shocks Congestion is guided by the parameter that governs the elasticity of substitution between worker types, \( \sigma \), which determines the degree of diminishing returns to specific worker types. We parameterize \( \sigma \) (jointly with the two aggregate shock processes above) by having the model match the impulse response of labor market tightness to a separation shock, estimated using the same VAR as in Section 2.2 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 8 plots, as the blue solid line, how this RMSE varies with the congestion parameter \( \sigma \). We obtain the best fit at \( \sigma = 0.241 \). 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 9 Panel (a) 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. 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 2.2. Crucially, the inability of the no-congestion model to match the IRF is not a matter of calibration. In Appendix I, 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.

\(^{21}\)We target UE flows as a share in employment because of their key role for our congestion channel. We discuss alternative calibration approaches below.
Figure 8: Calibrating Congestion Parameter $\sigma$

Notes: For various values of congestion parameter $\sigma$, 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.

Figure 9: Impulse Response Functions to a Separation Rate Shock: Data and Models

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.

Additionally, Panel (b) of Figure 9 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 (Panel (a)), which pushes down the job finding rate. This contrast reflects the intuitions conveyed by the benchmarks discussed earlier in Section 2.2. The reason why even the congestion model does not fully capture the persistence of the empirical unemployment response in the data is a lower
calibrated persistence of separation shocks in our model. This is because our parameterization strategy puts front and center UE flows, which are key to the congestion mechanism. As a result, however, separations are more volatile (reflected in the higher initial response in unemployment in Panel (b) of Figure 9) and less persistent compared to the data. In Appendix J, we present an alternative calibration which, instead, matches the cyclical patterns of EU flows. This alternative calibration matches the unemployment response more closely, but as a cost exhibits less congestion as the fluctuations of UE flows are underpredicted. We discuss these properties again below in Section 4.5 when studying overall unemployment fluctuations in the model.

Validation of the Productivity Channel and Alternative Calibration of \( \sigma \): The Excess Procyclicality of New Hires’ 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 cyclicalities of newly hired and average workers.

Figure 10 reiterates the structure of our main calibration Figure 8, 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 \( \sigma \) values. In particular, for each value of \( \sigma \), 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 hence have homogeneous productivity and wages. The semi-elasticity ratio is therefore one, depicted as the rightmost value of \( \sigma \).

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 cyclicalities observed in the US micro data, as reported by the canonical meta-analysis in Pissarides (2009) (Table II therein).²² 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 fluctuations.

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 parameter, 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’ve assumed 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.

²²Recall 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).
4.2 Robustness to Alternative Congestion Mechanisms

Here we show that alternative model structures yield similar amplification to our baseline model, as long as the parameters guiding congestion are recalibrated to match the empirical IRF of labor market tightness to separation shocks.

A Model Featuring Both Congestion Hires 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., not losing 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 \( \sigma \) parameter, and ultimately exhibit the same degree of congestion.
Figure 11: Robustness to Alternative Specifications of Skill Process

![Graph showing robustness to alternative specifications of skill process](image)

Notes: The figure plots recalibrated values of $\sigma$ 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 $\sigma$ is recalibrated to match the market-tightness impulse response target.

To demonstrate robustness, we detail an extreme alternative of the type evolution in Appendix K. 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. To achieve this fit, each $x$-model simply requires a lower and lower $\sigma$. We plot the resulting “iso-congestion” $\sigma(x)$ curve with the blue solid line. Appendix K derives this iso-congestion curve analytically.

Importantly, the dotted red lines show that the standard deviation of unemployment is invariant in $x$ along the iso-congestion curve for $\sigma$—so such alternative specifications of the skill process are isomorphic to our baseline specification in which all workers fall to $k = 1$ upon job loss. Appendix K provides further details and formally shows this isomorphism.

---

23 Here, the $a_k$-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 the congestion (CRS-CES) and a no-congestion (linear) production functions, $Y = z[(1 - x)\alpha^c_k n^c_k + x(\alpha^nc_k n^nc_k)]$, where subscripts $c$ and $nc$ stand for the congestion and no-congestion sectors.

24 There, 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 $n_1$: $\epsilon_{\bar{p}_1,n_1} = (\sigma - 1)(1 - n_1/N)(1 - x)$. The iso-congestion curve for a desired degree of congestion $\bar{\tau}$ as a function of no-congestion worker share $x$ is given by $\sigma(x, \bar{\tau}) = 1 + \frac{\bar{\tau}}{(1 - x)/(1 - n_1/N)}$. This analytical curve turns out to be essentially identical to the blue line.
Table 2: Business Cycle Properties: Data and Congestion Model

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
<th>p_1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.053</td>
<td>0.067</td>
<td>0.103</td>
<td>0.126</td>
<td>0.229</td>
<td>0.067</td>
<td>NA</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.746</td>
<td>0.871</td>
<td>0.773</td>
<td>0.934</td>
<td>0.926</td>
<td>0.936</td>
<td>0.836</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>0.042</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>-0.415</td>
<td>-0.715</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>-0.112</td>
<td>-0.931</td>
<td>0.848</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>0.309</td>
<td>0.874</td>
<td>-0.869</td>
<td>-0.934</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ</td>
<td>0.223</td>
<td>0.917</td>
<td>-0.874</td>
<td>-0.980</td>
<td>0.986</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UE/E</td>
<td>0.173</td>
<td>-0.722</td>
<td>0.567</td>
<td>0.833</td>
<td>-0.711</td>
<td>-0.783</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

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

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. Panel A reports values from the data; Panel B reports these values for our baseline model. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.

**Congestion Through Convex Hiring Costs.** In addition, in Appendix L 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. 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 DMP recruitment costs, 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 5.

See, e.g., Fujita and Ramey (2007); Coles and Moghaddasi Kelishomi (2018) for models that relax the free entry condition along those lines.
4.3 A Bird’s Eye View of Business Cycle Statistics

We now study the quantitative implications of countercyclical congestion for labor market fluctuations. Table 2 provides an overview of business cycle statistics for the data (Panel A) and our congestion model (Panel B). The model closely replicates the business cycle properties of the key empirical variables, both with regards to volatility and cyclicality. 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).

Robustness: Alternative Specifications. Appendix M replicates Table 2 for the no-congestion model, which is isomorphic to the standard DMP model calibrated and studied by Shimer (2005), reporting the variants with and without separation shocks. Appendix I does so for the no-congestion model under the Hagedorn and Manovskii (2008) calibration, i.e., featuring a small match surplus in steady state (high $b$ compared to productivity), which permits productivity shocks to have a larger effect on hiring and generate realistic labor market volatility (Ljungqvist and Sargent, 2017). Appendix L provides the model that obtains congestion through countercyclical UE hiring costs. We have additionally experimented with a model featuring decreasing returns in aggregate employment, similar to Michaillat (2012); that model variant would predict essentially no effect of separation rate shocks on labor market tightness, as in the standard DMP model, for lack of a congesting-in-hiring mechanism.

The congestion model does estimate higher volatility of the separation shock $\delta$ compared to the data, but with lower autocorrelation. This is because we calibrate the separation rate process to match the UE flows in the data rather than EU flows. We discuss this choice in Appendix J, where we additionally present an alternative calibration of the separation rate process to EU flows consistent with a two-state model. There, the separation rate has a lower standard deviation and higher autocorrelation. Matching this target comes at the cost of missing the UE flows, such that congestion is attenuated and as a result hiring fluctuations are less volatile (a standard deviation of 0.144 rather than 0.207). We suspect that a three-state model with an out of the labor force state would permit matching all targets.

4.4 Beveridge Curves

We now study the Beveridge curve, the relationship between vacancies and unemployment, and a crucial property of DMP models (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 12 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.

The no-congestion model features a counterfactually positive slope: 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 shocks, as described in Sections 2.2 and 4.1. 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. 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, exactly this process of expanding gross flows diminishes the returns to further hiring, permitting the model to rationalize elevated unemployment.²⁶

### 4.5 The Volatility of Unemployment

Figure 8 visualizes how congestion leads to amplification, by additionally plotting, with a red dashed line, the volatility of unemployment for different values of $\sigma$ (while recalibrating all other

---

²⁶Coles 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.
parameters to match the remaining targets). Consider a recession. As separations increase, unemployment rises. UE flows rise, which lowers their type-specific marginal product of labor, as long as $\sigma < 1$, so that their productivity is much more volatile and procyclical than average labor productivity (last column of Table 2 Panel B). This productivity drop further reduces hiring incentives, keeping unemployment elevated.

Importantly, the amplification and propagation relative to the standard DMP framework is exclusively due to countercyclical congestion, i.e., the degree to which shifts in the employment share of new hires diminish their productivity. We surgically isolate the congestion channel as we maintain the same, high fundamental match surplus for each model. We do so by recalibrating the productivity weights $\alpha_k$ to generate the common unit productivity instead state for all types, as described in Section 4.1.

4.6 Sources of Amplification: Productivity and Cohort Dynamics

The key to understanding amplification is the behavior of the match surplus for new hires. Using Equation (17) 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

$$S_{k,t} = p_{k,t} - b + \beta E_t \left[ (1 - \delta_{t+1}) S_{k+1,t+1} \right] - \beta E_t \left[ (1 - \delta_{t+1}) f(\theta_t) \phi S_{1,t+1} \right].$$

(24)

In comparison to the no-congestion model, amplification in surplus fluctuations stems from three sources. First, the flow productivity channel works through 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 (24) to explicitly highlight these three amplification channels, now specifically focusing on the surplus of new hires $k = 1$:

$$S_{1,t} = z_t - b + \beta E_t \left[ (1 - \delta_{t+1})(1 - f(\theta_t))S^s_{t+1} \right] + \beta E_t \left[ (1 - \delta_{t+1}) f(\theta_t)\phi S^s_{1,t+1} \right] = S^s_{t} + S^s_{t+1}$$

(i) No-congestion model surplus

$$+ \beta E_t \left[ (1 - \delta_{t+1})(1 - f(\theta_t)\phi) \left( S_{2,t+1} - S^s_{t+1} \right) \right] + \beta E_t \left[ (1 - \delta_{t+1}) f(\theta_t)\phi (S_{2,t+1} - S_{1,t+1}) \right].$$

(25)

(iii) Present value channel

(iv) Outside option channel

where $S^s_{t} = z_t - b + \beta E_t \left[ (1 - \delta_{t+1})(1 - f(\theta_t)\phi)S^s_{t+1} \right]$ is the surplus in the standard model without congestion and homogeneous workers, and $\theta^s$ is the associated labor market tightness.\(^{27}\)
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.

\[ S_t^* = p_{1,t} - b + \beta \mathbb{E}_t \left[ (1 - \delta_t) (1 - f(\theta_t) \phi) S_{t+1}^* \right] \] 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 7, 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. 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 13 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.\(^{28}\)

\(^{28}\)The 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.
As a result of these persistent cohort effects, the expected present value of productivity of newly hired workers (formally, $E_t \sum_{j=0}^{\infty} \beta^j (1-\delta_{t+1+j}) p_{1+j,t+1+j}$)—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.

**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, 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.29

**Quantifying the Sources of Amplification.** We now quantify the contributions of the three channels to amplification arising from countercyclical congestion. We do so by feeding in counterfactual surpluses from subsets of the four channels in Equation (25) into the free-entry condition in Equation (21). 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 percent of the baseline fluctuations. Therefore, the outside option channel explains only 15 percent 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 percent of the variation in labor market tightness. Finally, the no-congestion model accounts for only about 5 percent 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.

### 4.7 Historical Decomposition of Unemployment in the United States

We now study how countercyclical congestion has historically 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’

---

29This mechanism would not be present with wage setting protocols that insulate wages wage from outside options (Hall and Milgrom, 2008; Jäger, Schöfer, Young, and Zweimüller, 2020).
Table 3: Volatility of Labor Market Tightness and Sources of Amplification

<table>
<thead>
<tr>
<th>Source of Amplification</th>
<th>Standard Deviation</th>
<th>Contribution to Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-congestion model (i)</td>
<td>0.019</td>
<td>0.049</td>
</tr>
<tr>
<td>+ Flow productivity channel (i)+(ii)</td>
<td>0.052</td>
<td>0.162</td>
</tr>
<tr>
<td>+ Present value channel (i)+(ii)+(iii)</td>
<td>0.178</td>
<td>0.851</td>
</tr>
<tr>
<td>+ Outside option channel (i)+(ii)+(iii)+(iv)</td>
<td>0.207</td>
<td>1</td>
</tr>
</tbody>
</table>

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 \[\text{cov}(\vartheta_{\text{base}}, \vartheta_{\text{cf.}}) / \text{var}(\vartheta_{\text{base}})\], where \(\vartheta_{\text{base.}}\) is labor market tightness in our baseline model, while \(\vartheta_{\text{cf.}}\) is the respective counterfactual labor market tightness.

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:

\[
\begin{align*}
    u_{t+1}^c &= (1 - f(\vartheta_t^c))u_t^c + \bar{\delta}(1 - u_t^c) \\
    \kappa &= q(\vartheta_t^c)\beta_{t}E_t(1 - \bar{\delta})S_{t,t}^c \\
    S_{k,t}^c &= p_{k,t} \cdot \frac{z_t}{z_t} \cdot b + \beta_{t}E_t(1 - \bar{\delta})S_{k+1,t+1}^c - \beta_{t}E_t(1 - \bar{\delta})f(\vartheta_t^c)\phi S_{1,t+1}^c \quad \forall k
\end{align*}
\]

(26)

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 A17 presents both the estimated and empirical time series, Appendix G contains further details on the estimation procedure. Appendix N provides additional details on the decomposition, and additionally applies the method to TFP-only and separation-only counterfactuals.
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 (26).

**The Time Series of Congestion-Driven Unemployment.** Figure 14 shows the time path of congestion unemployment in the US, and compares it to overall unemployment (which, which model-implied, essentially perfectly tracks the empirical time series, as shown in Appendix Figure A17). 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 percent the level of of overall unemployment. Computing the contribution of congestion-only unemployment, we find $\text{cov}(u, u^c)/\text{var}(u) = 0.297$ (with a correlation of 0.723). Therefore, countercyclical congestion explains 30 to 40 percent of observed unemployment fluctuations.

5 **Additional Implications of Countercyclical Congestion**

Besides providing a new perspective on unemployment fluctuations, countercyclical congestion rationalizes three additional rekted 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 sensitivity of labor market variables to labor market policies. These applications provide additional, external validation.
5.1 Business Cycle Accounting: The Labor Wedge

The Standard Labor Wedge. In a perfectly competitive spot labor market with representative agents, as in real business cycle (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:

\[
(1 - \tau) \cdot MPL = MRS \left( = \frac{-U_E(C, E)}{U_C(C, E)} \right). \tag{27}
\]

Figure 15 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). As is well known, the US data exhibit a volatile and procyclical 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. Our 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 \( \eta \) 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).

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).

Congestion and the Labor Wedge. To show that the more procyclical marginal product of labor implied by our congestion model offers such an explanation, 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:

\[
Y = \tilde{z} \tilde{K}^a \cdot \left( \sum_{l=1}^{K} \alpha_l n_l^a \right)^{1-a}. \tag{28}
\]

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
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 (29). All series are in logs and HP filtered using a smoothing parameter of 1,600.

Figure 15 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). The remaining variation of the labor wedge after subtracting the new-hire term is essentially unrelated to the business cycle: the elasticity with respect to the detrended unemployment rate falls from −0.328 ($R^2 = 0.872$) to 0.081 ($R^2 = 0.111$). That is, the economy with congestion provides an essentially full explanation of the labor wedge.

5.2 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 study complements our study of new hires’ flow wages in Section 4.1.

We construct the term as the new-hire productivity in Figure 7 Panel (b), described in Footnote 15.
The Cyclicality of Displacement Costs in the Congestion Model. A large body of research has documented large and persistent earnings losses following job displacement events, of around 30 percent 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 countercyclicality of earnings losses from displacement. We replicate the analysis in Davis and von Wachter (2011). We compute the earnings trajectory of a cohort of separated workers, taking 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 generating the average 3.5 percentage point unemployment rate difference between troughs and peaks of NBER-dated business cycles 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 of 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 16 Panel (a) shows the difference in earnings effects from a job separation in recessions compared to booms for the model (blue solid line). 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. The empirical earnings losses are about 13 percentage points larger in recessions compared to booms. After 10 years, this difference declines to about 6 percentage points. Our model features flexible wages (Nash bargaining with a high bargaining power of workers), which are therefore quite sensitive to match-specific productivity and may, therefore, lead it to predict higher displacement effects in the middle years.

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

---

31Since our model is calibrated such that all worker types have identical wages in steady state ($p_k = 1$ and hence $w_k = w_k$ for all $k$), it cannot speak to the level of displacement costs.

32The 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 16 Panel (a).
von Wachter (2019):

\[ y_{g,t} = \alpha + \beta_e u_g + \lambda_g + \chi_t + \epsilon_{g,t}, \]  

(30)

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”), \( \lambda_g \) are graduation fixed effects, and \( \chi_t \) are time fixed effects. The coefficients of interest are given by vector \( \beta_e \), which captures the effect of the unemployment rate at the time of labor market entry on subsequent earnings, where \( \epsilon = t - g \) captures time since graduation.

Figure 16 Panel (b) plots the \( \beta_e \) coefficients estimated on simulated data together with the empirical estimates from Schwandt and von Wachter (2019). The model closely matches the data, with a one percentage point increase in unemployment resulting in about a 3.5 percent 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. Most of the proximate sources of these two types of countercyclical earnings losses are accounted for by wage declines conditional on working (see, e.g., von Wachter, forthcoming), supporting the persistent cohort effects on productivity in our model. The cohort effects also are 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). This pattern could be viewed as consistent with congestion manifesting itself as low-quality relative to high-quality firms absorbing the increase in UE hires. A complementary literature studies the destruction and creation of jobs by firm qual-

---

33See Figure 2 in Schwandt and von Wachter (2019) for the empirical estimates in our Figure 16 Panel (b).
ity (Moscarini and Postel-Vinay, 2012; Haltiwanger, Hyatt, Kahn, and McEntarfer, 2018), and the countercyclicality of skill requirements (Modestino, Shoag, and Ballance, 2016).

5.3 Policy Insensitivity Despite Productivity-Driven Business Cycles

We close by revisiting 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 sensitivity of these variables with respect to policies that affect job surplus, such as unemployment insurance (UI) benefits. Costain and Reiter (2008) estimate the semi-elasticity of the unemployment rate with respect to the replacement rate, \( \epsilon_{u,b/w} = \frac{\partial \ln u}{\partial (b/w)} \), to lie between 2 and 3.5 in a cross-country analysis. While the standard DMP model can replicate this semi-elasticity, it fails to generate sufficient volatility in labor market variables. By contrast, the solution by Hagedorn and Manovskii (2008) to calibrate steady state \( b \) to feature a small fundamental surplus (Ljungqvist and Sargent, 2017), generates sufficient volatility in labor market variables, but overstates the sensitivity to UI (for empirical research on short run effects across US local labor markets, see Hagedorn, Karahan, Manovskii, and Mitman, 2019; Chodorow-Reich, Coglianese, and Karabarbounis, 2019; Boone, Dube, Goodman, and Kaplan, forthcoming).

Returning to our model and starting from our baseline calibration, we increase the UI benefit level \( b \) by 1 percent, i.e., \( b_{\text{new}} = 1.01b_{\text{base}} \), and recompute the steady state values for all the model variables. Following Costain and Reiter (2008), we then compute the semi-elasticity of unemployment with respect to the replacement rate as \( \epsilon_{u,b/w} = \frac{\ln u_{\text{new}} - \ln u_{\text{base}}}{(b_{\text{new}}/w_{\text{new}}) - (b_{\text{base}}/w_{\text{base}})} \approx 2.6 \), 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, so it can afford small elasticities.

6 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 these new hires, countercyclical UE flows can generate a mechanism we call countercyclical congestion. Due to diminishing returns in the types of jobs the unemployed fill, the labor productivity of new hires falls by much more than average labor productivity, lowering further hiring incentives, and raising unemployment.

The model with countercyclical congestion is consistent with a range of macroeconomic regularities. In particular 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 the countercyclical labor wedge, the excess procyclicality of new hires’ wages, the countercyclical earnings losses upon job displacement and labor market entry.

We close with questions our study leaves open. First, we have presented aggregate time series
evidence consistent with congestion and reviewed cross sectional quasi-experimental studies—but we have not definitively quantified the degree of congestion. Second, our study suggests that factors and policies attenuating shifts in separations, such as firing taxes or furlough schemes, may also attenuate shifts in the job finding rate. Third, while our collage of wage-based evidence has supported our productivity-based congestion mechanism, we have shown that congestion may emerge also from hiring costs or perhaps other factors.
References


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

Contents

A Measuring Worker Flows .......................... 51
B Evidence from OECD Countries ................. 57
C Discrete versus Time-Aggregation-Adjusted Data 59
D Alternative HP Smoothing Parameter .......... 64
E Robustness of Separation Shock Identification to Other Macroeconomic Shocks 66
F A Generalization of the Baseline Model: Types vs. Inputs 69
G Solution Method .................................. 70
H Details of the Baseline Parameterization: Homogeneous Steady State Marginal Products Across Types 72
I Alternative Calibration: Small Surplus/“High b” 73
J Alternative Calibration: Accounting for Non-Participation and Matching EU Flows 76
K Deriving the Iso-congestion Curve ............... 81
L Alternative Mechanism: Convex Hiring Costs 84
M Business Cycle Statistics in the Standard Model 86
N Historical Decomposition: Additional Material 88
List of Tables

A1  Discrete vs. Time-Aggregation Adjusted Worker Transitions .............................................. 60
A2  Business Cycle Properties: Alternative Smoothing Parameter ............................................ 64
A3  Separation Shocks and Other Disturbances: Adjusted R-squared ....................................... 67
A4  Business Cycle Properties: No-Congestion, Low-Surplus Model ....................................... 75
A5  Business Cycle Properties of the Congestion Model: Data and Extended Model for NU Flows ..................................................................................................................... 79
A6  Business Cycle Properties: Convex Hiring Cost Model ......................................................... 85
A7  Business Cycle Properties of the No-Congestion Model ....................................................... 87
A8  Historical Decomposition of Unemployment: Model and Counterfactuals ............................ 90

List of Figures

A1  Countercyclicality of the Employment Share with Nonemployment Past Year ....................... 53
A2  The Countercyclicality of New Hire Share: CPS Worker Flows ............................................ 54
A3  Cyclicality of Share of New Hires in Employment: CPS Worker Flows ............................... 55
A4  Cyclicality of New Hires: CPS Worker Flows ......................................................................... 56
A5  Cyclicality of UE Flows in the OECD ....................................................................................... 58
A6  Comparing Discrete and Time Aggregation Adjusted UE Shares .......................................... 61
A7  Comparing Discrete and Time Aggregation Adjusted Flow Probabilities ............................. 63
A8  The Countercyclicality of Unemployment-to-Employment Flows .......................................... 65
A9  Separation Shocks: Baseline and Adjusted for Other Disturbances ..................................... 68
A10劳 Productivity: Empirical Impulse Responses to a Separation Shock .................................. 71
A11 Relative Worker Productivities in the Congestion Model ...................................................... 72
A12 Impulse Responses to a Separation Shock: No-Congestion, Low-Surplus Model .................. 74
A13 Separation Rate: Measured and Implied .................................................................................. 77
A14 Impulse Responses to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows ................................................................. 80
A15 Iso-congestion Curves ............................................................................................................. 83
A16 Impulse Responses to a Separation Shock: Convex Hiring Cost Model ............................... 85
A17 Time Paths of Labor Market Variables ................................................................................... 89
A18 Unemployment Components ................................................................................................ 90
A Measuring Worker Flows

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. The matching rate is below 75 percent due to temporary absence of individuals in 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.

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.

As a robustness check discussed in Section 2.1, below in Figures A1 to A4, we also replicate main text 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 in order to match our empirical calibration targets, which we describe in Section 2.2, with the model calibration strategy described in Section 4.1.
Figure A1: Countercyclicality of the Employment Share with Nonemployment Past Year

(a) Employment Shares of Workers with Nonemployment Last Year by Total Weeks

(b) Cyclicality: Log Deviations from Trend

(c) Okun’s Law

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).
Figure A2: The Countercyclicality of New Hire Share: CPS Worker Flows

(a) UE Share in Employed

(b) NE Share in Employed

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.
Figure A3: Cyclicality of Share of New Hires in Employment: CPS Worker Flows

(a) UE Share vs. Unemployment Rate

(b) NE Share vs. Unemployment Rate

(c) UE Share vs. E-Population Ratio

(d) NE Share vs. E-Population Ratio

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.
Figure A4: Cyclicality of New Hires: CPS Worker Flows

(a) UE Flows vs. Unemployment Rate

(b) NE Flows vs. Unemployment Rate

(c) UE Flows vs. E-population Ratio

(d) NE Flows vs. E-population Ratio

Notes: This figure is a complement to Figure A3. 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 A1. Source: CPS monthly files.
B Evidence from OECD Countries

The countercyclicality of UE flows extends to many OECD countries. In Figure A5 Panel (a), we plot the elasticity of UE flows with respect to the unemployment rate for a set of 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 in Equation (4), building on the insight that the unemployment rate fluctuations implied by the job finding rate shift only is $\frac{du^f}{u^f} = -(1 - u)\frac{df}{f}$. Fujita and Ramey (2009) show that the regression coefficient of $\frac{du^f}{u^f}$ on $\frac{du}{u}$ 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 $\frac{dUE}{UE} \frac{du}{u} = 1 - \frac{du^f}{u^f} \frac{du}{u} + 1$. Drawing on cross-country differences in the OECD, we document the empirical validity of this theoretical property in Panel (b) of Figure A5, a scatterplot that shows a clear negative relationship between the elasticity against the contribution of job finding rate to unemployment fluctuations, the latter 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 implied by Equation (4). Broadly, the relationship between the two elasticities holds across countries (with the approximation error reflecting the assumptions of steady state and two states).
Figure A5: Cyclicality 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 $\text{cov}(-\hat{u}_{ss}(1-u_{ss}), \hat{f} u_{ss})/\text{var}(u_{ss})$, 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 $-\hat{u}_{ss}(1-u_{ss})\hat{f}$ is the unemployment rate deviation due to the job finding rate only. For the DMP model without separations shocks, this share is one, and the elasticity on the y-axis is computed using formula (4). 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.
C  Discrete versus Time-Aggregation-Adjusted Data

Our preferred measure of worker flows used in the main text is based on discrete time and hence 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 appendix, 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 remain active at the end of the period, mirroring our definition using the CPS ASEC in Section 2.1. We also confirm that our time series replicate to those reported by Shimer (2012).

Our Method. Our method draws on Fujita and Ramey (2006), who provide expressions for time-aggregation-adjusted gross worker flows, whereas our interest is in cumulative UE flows that remain active through the end of the period. We describe our method below. With an abuse of notation, we denote the discrete job finding and loss probabilities by \( \hat{f} \) and \( \hat{\delta} \) (which correspond to \( f \) and \( \delta \) in the main text).

We start with the monthly job finding \( \hat{f}_t \) and separation \( \hat{\delta}_t \) probabilities, whose measurement are described in Appendix A, underlying the analysis in the main text.

Second, we compute the monthly job finding and separation hazards, \( f_t \) and \( \delta_t \), solving the following system of equations:

\[
\hat{\delta}_t = u_{ss,t}(1 - e^{-f_t - \delta_t}) \\
\hat{f}_t = (1 - u_{ss,t})(1 - e^{-f_t - \delta_t}),
\]

where \( u_{ss,t} = \delta_t/(\delta_t + f_t) \) is the steady-state approximation to the unemployment rate implied by the contemporaneous transition rates. The law of motion for unemployment in continuous time is given by

\[
U_{t-1+\tau} = \frac{(1 - e^{-(f_t+\delta_t)\tau})\delta_t}{f_t + \delta_t} L_{t-1} + e^{-(f_t+\delta_t)\tau} U_{t-1},
\]

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

<table>
<thead>
<tr>
<th></th>
<th>Discrete</th>
<th>Time-aggregation adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.045</td>
<td>0.040</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.671</td>
<td>0.574</td>
</tr>
</tbody>
</table>

Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Discrete</th>
<th>Time-aggregation adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.983</td>
</tr>
</tbody>
</table>

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.

\( \tau \in [0, 1) \), conditional on having a job at \( t \). This probability is given by

\[
\lim_{\Delta \to 0} (1 - \Delta \delta_t)^{\frac{1-\tau}{\Delta}} = e^{-\delta_t(1-\tau)}. \tag{A3}
\]

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

\[
UE_t = \int_0^1 f_t \left( \frac{U_{t-1+\tau}}{U_{t-1} - U_{t-1,\tau}} \right) \frac{e^{-\delta_t(1-\tau)}}{f_t} d\tau. \tag{A4}
\]

Finally, using Equation (A2), we can integrate out the above expression to obtain UE flows adjusted for time aggregation bias:

\[
UE_t = f_t L_{t-1} e^{-\delta_t} \left( u_{\text{ss},t} e^\delta_t - 1 \right) \frac{U_{t-1}}{L_{t-1}} - u_{\text{ss},t} \left( 1 - e^{-f_t} \right). \tag{A5}
\]

Table A1 summarizes the properties of the time series we use in the main text and the time series we construct using the alternative approach presented above. The two time series have extremely similar standard deviations and autocorrelations, and are nearly perfectly correlated.

Figure A6 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 2 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, \( f \) and \( \delta \), in Panel (c) of

---

1Therefore, our results do not study cycles such as “EUEUEUE” 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.
Figure A6: Comparing Discrete and Time Aggregation Adjusted UE Shares

(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, $f$ and $\delta$. 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.

Figure A6, 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). As further supporting evidence pointing at the credibility of our measurement exercise and validity of our empirical analysis, we compare our preferred worker transition probabilities to the ones reported in Shimer (2012). Panel (a) in Figure A7 plots the employment-to-unemployment probability used in the main text and compares that to the monthly
probability adjusted for time aggregation using the method in Shimer (2012) allowing for flows in and out of inactivity (i.e. labor force participation). 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) data would be similarly countercyclical.²

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 cyclicality” (page 145). Our reading is that this statement likely assesses the magnitude of 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 A7: Comparing Discrete and Time Aggregation Adjusted Flow Probabilities

(a) EU Probabilities

(b) UE Probabilities

(c) Cyclical EU Probabilities

(d) Cyclical UE 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 1600. Source: CPS monthly files.
D Alternative HP Smoothing Parameter

In the main text, we report business cycle statistics based on HP-filtered time series with a smoothing parameter of 1600, 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 $f$, $\delta$ and $UE/E$ become if anything even more pronounced.\(^3\)

Table A2: Business Cycle Properties: Alternative Smoothing Parameter

<table>
<thead>
<tr>
<th></th>
<th>$ALP$</th>
<th>$f$</th>
<th>$\delta$</th>
<th>$u$</th>
<th>$v$</th>
<th>$\theta$</th>
<th>$UE/E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.017</td>
<td>0.093</td>
<td>0.108</td>
<td>0.190</td>
<td>0.198</td>
<td>0.376</td>
<td>0.116</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.897</td>
<td>0.950</td>
<td>0.904</td>
<td>0.970</td>
<td>0.957</td>
<td>0.962</td>
<td>0.933</td>
</tr>
</tbody>
</table>

Correlation matrix

\[
\begin{array}{ccccccc}
ALP & 1 & & & & & \\
\hline
f & -0.061 & 1 \\
\delta & -0.179 & -0.859 & 1 \\
u & 0.015 & -0.975 & 0.919 & 1 \\
v & 0.050 & 0.831 & -0.830 & -0.851 & 1 \\
\theta & 0.038 & 0.906 & -0.877 & -0.928 & 0.978 & 1 \\
UE/E & 0.113 & -0.888 & 0.783 & 0.930 & -0.718 & -0.818 & 1 \\
\end{array}
\]

Notes: $ALP$, $f$, $\delta$, $u$, $\theta$ 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.

Most importantly, Figure A8 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 the main text in Figure 2 Panel (b) (0.348 vs 0.345). Likewise, the elasticity of new-hire share in employment to the unemployment rate stays unchanged compared to the one reported in Figure A3 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).

\(^3\)The 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 A8: The Countercyclicality of Unemployment-to-Employment Flows

(a) Unemployment vs. UE Flows

(b) Unemployment vs. UE Share

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 $10^3$. Source: CPS monthly files.
E Robustness of Separation Shock Identification to Other Macroeconomic Shocks

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 $\sigma$, which governs the extent of congestion.

In this appendix, we show that our identified separation shocks are indeed only marginally affected by other identified structural shocks considered in the literature. Specifically, we consider total factor productivity shocks (Fernald, 2014), financial shocks (Gilchrist and Zakrajšek, 2012), discount factor shocks (Hall, 2017), uncertainty shocks (Jurado, Ludvigson, and Ng, 2015) and monetary policy shocks (Romer and Romer, 2004; Wieland and Yang, 2020).

E.1 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.

**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.

---

4 An 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, the available shock series cover a considerably shorter sample period.
Table A3: Separation Shocks and Other Disturbances: Adjusted R-squared

<table>
<thead>
<tr>
<th></th>
<th>TFP</th>
<th>Financial</th>
<th>Discounts</th>
<th>Uncertainty</th>
<th>Monetary Policy</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>-0.006</td>
<td>0.004</td>
<td>0.016</td>
<td>-0.007</td>
<td>-0.018</td>
<td>-0.096</td>
</tr>
<tr>
<td># of obs.</td>
<td>156</td>
<td>119</td>
<td>138</td>
<td>156</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td># of coefs.</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>26</td>
</tr>
</tbody>
</table>

Notes: The top row reports the adjusted R-square from the individual regressions (A6) for the five different macroeconomic shocks and the “joint” regression in Equation (A7). “TFP” is the utilization-adjusted total factor productivity shock from 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 et al. (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.

E.2 Separation Shocks and Other Macroeconomic Disturbances

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

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

(A6)

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

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

(A7)

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

Table A3 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 A9 shows how the separation shocks estimated in the main text change when controlling for all the above macroeconomic shocks using the regression model in Equation (A7). The figure reveals that the estimated shocks are largely unchanged, as suggested by the slightly negative $R^2$ in Table A3.
Figure A9: Separation Shocks: Baseline and Adjusted for Other Disturbances

Notes: The figure shows the baseline separation shocks estimated in Section 2 and those shocks “adjusted for other disturbances” using the regression mode in Equation (A7), where the plotted series is $\eta_t$. 

log deviation from steady state

Baseline

Adjusted for other shocks

F  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., \( Y = z \left( \sum_{k=1}^{K} a_k n_k^\sigma \right)^{\frac{1}{\sigma}} \). In this Appendix we generalize this setup by allowing for subsets of worker types $i \subset K$ 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 $i \in I = \{1, \ldots, I\}$ is defined by a set of worker types $\Omega_i \subset K$ which are mutually exclusive, i.e., \( \cap_i \Omega_i = \emptyset \). The production function in this setting is given by \( Y = z \left( \sum_i a_i n_i^\sigma \right)^{\frac{1}{\sigma}} \).

This setup of worker heterogeneity nests multiple cases. For example, if $I = 1$, then $\Omega_1 = K$ 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 $\Omega_1 = \{1, \ldots, 12\}$ and $\Omega_2 = \{13, \ldots, K\}$. 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 $i = 1, \ldots, K$.

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:

\[
\begin{align*}
    p_i &= \alpha_i n_i^{\sigma-1} \frac{Y}{\sum_j \alpha_j n_j^{\sigma-1}} = \alpha_i s_i^{\sigma-1} \frac{1}{\sum_j \alpha_j s_j^{\sigma-1}} Y \frac{1}{N},
\end{align*}
\]

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

Here, also the worker and firm values now reflect the fact that worker types, $k$, 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 $k$. 

[69]
G 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 a distribution of worker types in employment (and unemployment). We then lay out the solution method for the dynamic model and for its estimation.

G.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:

$$
e_{1} = fu,
\nonumber$$

$$
e_{k+1} = e_{k}(1-\delta) \text{ for } k = 1, ..., K-1,
\nonumber$$

$$
u = (1-f)u + \delta \sum_{k} e_{k}.
\nonumber$$

In addition, under our assumption 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 (21). 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 compute the implied productivity weights $a_{k}$ via

$$
1 = p_{k} = a_{k}s_{k}^{-\sigma-1}\frac{1}{\sum_{l=1}^{K} a_{l}s_{l}^{-\sigma} N},
\nonumber$$

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

G.2 Solution and Estimation with Aggregate Uncertainty

Our model features heterogeneity in worker types and two aggregate sources of heterogeneity, $z$ and $\delta$. The employment distribution gives another set of 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 (11), 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
Figure A10: Labor Productivity: Empirical Impulse Responses to a Separation Shock

(a) ALP: VAR including Market Tightness
(b) ALP: VAR including Unemployment

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 (8) with market tightness as the last variable. Panel (b) plots the impulse response of ALP in the VAR model with unemployment rate as the last variable. The separation shocks are identified off a Cholesky decomposition as explained in Section 2.1. The model IRFs exhibit a tiny increase initially in ALP and then a persistent but very small negative productivity effect for ALP; specifically, it is present for both the $\sigma = 1$ and $\sigma = 0.241$ models, yet it will not generate any noticeable reduction in labor market tightness for the latter economy (see the red dashed line in Figure 9), including in the small-surplus variant of the no-congestion model (Appendix Figure A12).

variables (the two aggregate shocks, the distribution of employment levels 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.

G.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 4.7 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 4.7 to calculate the contribution of congestion unemployment to the variation in observed unemployment fluctuations. Figure A17 shows the time paths of other labor market variables implied by our estimation.
H 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 $\alpha_k$ 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 A11 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 $\alpha_1$. 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 A11: Relative Worker Productivities in the Congestion Model](image)

Notes: The figure plots the relative productivity weights in production, $\alpha_k$, in the congestion model with $\sigma = 0.241$. The spike at $k = K$ ($= 160$) reflects the fact that it is an absorbing state.
I 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 A4. 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 2.1: 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 countercyclicality 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 shocks we discussed in Appendix M above.

Figure A12 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

\[
    r + \delta = \frac{(z - b)(1 - \phi)q(\theta)}{\kappa} - \phi f(\theta),
\]

where \( r \) is the interest rate such that \( \beta = 1/(1 + r) \). Taking \( z \) as given and Implicitly differentiating Equation (A9) with respect to \( \delta \) and \( \theta \) gives

\[
    d\delta = \frac{(z - b)(1 - \phi)q'(\theta)}{\kappa}d\theta - \phi f'(\theta)d\theta
    = -[\mu(r + \delta) + \phi f(\theta)]d\theta/\theta.
\]
Rearranging the above, we can then write the elasticity of $\theta$ with respect to $\delta$ as

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

(A11)

where $\gamma_{\text{Nash}} = \frac{r + \delta + \phi f(\theta)}{\mu(r + \delta) + \phi f(\theta)}$ 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 $\gamma_{\text{Nash}} \approx 1$. Moreover, these calibrations also result in the denominator in (A11) 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 $-2\delta$. 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 A4: Business Cycle Properties: No-Congestion, Low-Surplus Model

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Low fundamental surplus model: no δ shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.064</td>
<td>0</td>
<td>0.052</td>
<td>0.199</td>
<td>0.230</td>
<td>0.049</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.706</td>
<td>0.706</td>
<td>0</td>
<td>0.844</td>
<td>0.596</td>
<td>0.706</td>
<td>0.311</td>
</tr>
<tr>
<td><strong>Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ALP</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>f</strong></td>
<td>0.999</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>δ</strong></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>u</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.647</td>
<td>-0.648</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>v</strong></td>
<td>0.980</td>
<td>0.981</td>
<td>-</td>
<td>-0.486</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>θ</strong></td>
<td>0.999</td>
<td>1.000</td>
<td>-</td>
<td>-0.648</td>
<td>0.981</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>UE/E</strong></td>
<td>0.476</td>
<td>0.476</td>
<td>-</td>
<td>-0.270</td>
<td>0.477</td>
<td>0.476</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B: Low fundamental surplus model: with δ shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.064</td>
<td>0.082</td>
<td>0.090</td>
<td>0.177</td>
<td>0.227</td>
<td>0.068</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.691</td>
<td>0.689</td>
<td>0.560</td>
<td>0.825</td>
<td>0.558</td>
<td>0.689</td>
<td>0.623</td>
</tr>
<tr>
<td><strong>Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ALP</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>f</strong></td>
<td>0.999</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>δ</strong></td>
<td></td>
<td>1</td>
<td>-0.413</td>
<td>-0.430</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>u</strong></td>
<td></td>
<td></td>
<td>-0.674</td>
<td>-0.684</td>
<td>0.699</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>v</strong></td>
<td>0.933</td>
<td>0.929</td>
<td>-0.197</td>
<td>-0.368</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>θ</strong></td>
<td>0.999</td>
<td>1.000</td>
<td>-0.430</td>
<td>-0.684</td>
<td>0.929</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>UE/E</strong></td>
<td>0.005</td>
<td>-0.001</td>
<td>0.266</td>
<td>0.455</td>
<td>0.229</td>
<td>-0.001</td>
<td>1</td>
</tr>
</tbody>
</table>

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. Panel A reports values from the data, Panel B reports the same for our baseline model. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1,600.
Alternative Calibration: Accounting for Non-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 rates $\delta$.

In this Appendix, we show that this inability to match both realistic UE hires’ 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, the two 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 et al. (2015a) for the importance of the nonparticipation margin over the business cycle.

J.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 $f$, $\delta$ and $u$:

\[ u_{t+1} = (1 - f_t)u_t + \delta_{t+1}(1 - u_t). \] (A12)

However, it is possible to compute an implied empirical measure of EU separation rates $\delta_{t+1}^{imp}$ 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. Specifically, we compute this implied process as

\[ \delta_{t+1}^{imp} = \frac{u_{t+1} - (1 - f_t)u_t}{1 - u_t}. \] (A13)

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 $\delta_{t+1}^{imp}$ with the actual EU separation rate $\delta$ permits a useful diagnostic: whenever $\delta_{t+1}^{imp}$ exceeds $\delta$, 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 ($\delta$) and UE transitions ($f$)—where these have been

\footnote{This 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 A13: Separation Rate: Measured and Implied

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

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.

In what follows, we recalibrate the baseline model to account for the discrepancy described above. As will become clear, subject to a recalibration of our key parameter, $\sigma$, the extended model delivers essentially the same quantitative results while, at the same time, matching the observed variation in $EU$ flows.

### J.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 $k = 1$ (while retaining our assumption that all unemployed fall into $k = 1$) is given by

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

(A14)

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:

\[
OU_t = \rho_{OU} OU_{t-1} + \epsilon_{OU,t},
\]

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

### J.3 Parameterized Model with Empirical \(\delta\) 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 \(\delta\) to match the cyclical pattern of \(EU\) flows in the data.

In addition, we set \(\rho_{OU}\) and \(\sigma_{OU}\) to match the persistence and volatility of \(OU\) flows as a share of the labor force observed in the data.\(^6\) In addition, we allow for a correlation between \(\epsilon_\delta\) and \(\epsilon_{OU}\) to match the observed correlation between \(OU/(E+U)\) and the unemployment rate, which is 0.72.

Table A5 shows the business cycle statistics of the extended model and, for comparability, those observed in the data. The extended model matches not only the volatility of average labor productivity, but now also that of \(EU\) flows—specifically, the \(\delta\) 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 the match the impulse response of labor market tightness to separation shocks, the extended model requires a \(\sigma\) of 0.08. Under this calibration, however, the extended model delivers essentially identical dynamics as the baseline model, see Figure ??.

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 parameterized to match those observed in the data, has essentially identical amplification properties regarding labor market tightness and unemployment as the model 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).

\(^6\)Because 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 1.2%, with persistence of 0.57 and standard deviation of 0.001.
Table A5: Business Cycle Properties of the Congestion Model: Data and Extended Model for NU Flows

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>$f$</th>
<th>$\delta$</th>
<th>$u$</th>
<th>$v$</th>
<th>$\theta$</th>
<th>$UE/E$</th>
<th>$p_1$</th>
<th>$\theta$</th>
<th>$UE/E$</th>
<th>$p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Data</td>
<td>0.010</td>
<td>0.053</td>
<td>0.067</td>
<td>0.103</td>
<td>0.126</td>
<td>0.229</td>
<td>0.067</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Autocorrelation</strong></td>
<td>0.746</td>
<td>0.871</td>
<td>0.773</td>
<td>0.934</td>
<td>0.926</td>
<td>0.936</td>
<td>0.836</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$</td>
<td>0.042</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.415</td>
<td>-0.715</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u$</td>
<td>-0.112</td>
<td>-0.931</td>
<td>0.848</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v$</td>
<td>0.309</td>
<td>0.874</td>
<td>-0.869</td>
<td>-0.934</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.223</td>
<td>0.917</td>
<td>-0.874</td>
<td>-0.980</td>
<td>0.986</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UE/E$</td>
<td>0.173</td>
<td>-0.722</td>
<td>0.567</td>
<td>0.833</td>
<td>-0.711</td>
<td>-0.783</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: Congestion Model with OU Flows</td>
<td>0.010</td>
<td>0.054</td>
<td>0.067</td>
<td>0.099</td>
<td>0.099</td>
<td>0.189</td>
<td>0.052</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Autocorrelation</strong></td>
<td>0.701</td>
<td>0.901</td>
<td>0.544</td>
<td>0.850</td>
<td>0.889</td>
<td>0.902</td>
<td>0.767</td>
<td>0.781</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correlation matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$</td>
<td>0.323</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.419</td>
<td>-0.488</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u$</td>
<td>-0.337</td>
<td>-0.941</td>
<td>0.693</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v$</td>
<td>0.284</td>
<td>0.960</td>
<td>-0.241</td>
<td>-0.819</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.326</td>
<td>0.997</td>
<td>-0.491</td>
<td>-0.954</td>
<td>0.952</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UE/E$</td>
<td>-0.240</td>
<td>-0.938</td>
<td>0.390</td>
<td>0.890</td>
<td>-0.913</td>
<td>-0.946</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.414</td>
<td>0.950</td>
<td>-0.443</td>
<td>-0.882</td>
<td>0.926</td>
<td>0.948</td>
<td>-0.973</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: ALP, $f$, $\delta$, $u$, $\theta$, $UE/E$ 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. Panel A reports values from the data; Panel B reports the values for the model extended to include exogenous flows in and out of non-participation. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1,600.
Figure A14: 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.
K Deriving the Iso-congestion Curve

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

\[ Y = (1 - x) \left( \sum_k \alpha^c_k n^a_k \right)^{1/\sigma} + x \left( \sum_k \alpha^{nc}_k n_k \right). \]

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-\( k \):

\[ p_k = \frac{\partial Y}{\partial n_k} = (1 - x) \alpha^c_k n^{a-1}_k \left( \sum_k \alpha^c_k n^a_k \right)^{1/\sigma-1} + x \alpha^{nc}_k. \]

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 \( k \), \( \varepsilon_{p_k, n_k} \).

First, we observe that the elasticity of \( p^{nc}_k \) with respect to \( n_k \) is zero, \( \varepsilon_{p^{nc}_k, n_k} = 0 \). Second, we calculate the elasticity of \( p^c_k \) with respect to \( n_k \)

\[ p^c_k = \alpha^c_k n^{a-1}_k \left( \sum_k \alpha^c_k n^a_k \right)^{1/\sigma-1}, \]

\[ \Rightarrow \varepsilon_{p^c_k, n_k} = \frac{\partial p^c_k}{\partial n_k} \frac{n_k}{p^c_k} = (\sigma - 1) \left( 1 - \frac{\alpha^c_k n^a_k}{\sum_k \alpha^c_k n^a_k} \right). \]

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 \( \varepsilon_{p^{nc}_k, n_k} = 0 \), we derive our desired elasticity of marginal
product with respect to worker mass:

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

**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

\[ \varepsilon_{p_k, n_k} = (1 - x)(\sigma - 1) \left( 1 - \frac{\alpha_k n_k^\sigma}{\sum_k \alpha_k n_k^\sigma} \right). \] (A16)

Further, as \( p_k^c = \alpha_k n_k^\sigma\left( \sum_k \alpha_k n_k^\sigma \right)^{1/\sigma - 1} = 1 \) for all \( k \), we have \( \alpha_k n_k^\sigma = \alpha_1 n_1^\sigma \). Summing over \( k \), we get \( \sum_k \alpha_k n_k^\sigma = \alpha_1 n_1^\sigma N / n_1 \). Then we obtain \( s_1 = \frac{n_1}{N} = \frac{\alpha_1 n_1^\sigma}{\sum_k \alpha_k n_k^\sigma} \).

Using this result in the elasticity expression above, we finally arrive at

\[ \varepsilon_{p_k, n_k} = (1 - x)(\sigma - 1)(1 - s_k). \] (A17)

To trace out the iso-congestion curve for \( k = 1 \), we solve for \( \sigma \) as a function of \( x \) given a level of elasticity \( \varepsilon_{p_1, n_1} \).

\[ \sigma(x) = 1 + \frac{\varepsilon_{p_1, n_1}}{(1 - x)(1 - s_1)}. \] (A18)

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 \( \sigma \).

Figure A15 Panel (a) plots the iso-congestion curve derived in Equation (A18) 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, \( \sigma \) 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 9 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 A15: Iso-congestion Curves

(a) The Analytical Iso-congestion Curve

(b) Iso-congestion: Analytical vs. Model

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 9 Panel (a).
L 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:

\[ c(UE_t) = c_1 \cdot \left[ \left( \frac{UE_t}{UE_{ss}} \right)^{c_2} - 1 \right]. \]  \hspace{1cm} (A19)

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

\[ \frac{\kappa}{q_t} + c(UE_{t+1}) = E_t \left[ \beta(1 - \delta_{t+1})J_{t+1} \right]. \]  \hspace{1cm} (A20)

In turn, we remove worker heterogeneity (essentially setting \(\sigma = 1\) and setting the \(\alpha_k\)'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 \(\sigma\), we now set \(c_2\) such that the model minimizes the RMSE of the response of labor market tightness to separation shocks. Figure A16 shows that the fit of this model is excellent too, closely mirroring the IRF of our main specification in Figure 9. The estimated level of \(c_2\) is 1.2.

The results are presented in Table A6. 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. (Of course, the model with convex hiring costs would not generate cyclical displacement costs that are persistent, for lack of cohort effects.)

\[^7\text{Pissarides (2009); Silva and Toledo (2013) add a fixed costs of hiring, but it is not increasing in the amount of hires.}\]
Figure A16: Impulse Responses to a Separation Shock: Convex Hiring Cost Model

(a) Market Tightness

(b) Unemployment

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 A6: Business Cycle Properties: Convex Hiring Cost Model

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.061</td>
<td>0.118</td>
<td>0.129</td>
<td>0.096</td>
<td>0.219</td>
<td>0.067</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.691</td>
<td>0.855</td>
<td>0.536</td>
<td>0.845</td>
<td>0.856</td>
<td>0.855</td>
<td>0.840</td>
</tr>
</tbody>
</table>

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>0.505</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>−0.410</td>
<td>−0.726</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>−0.474</td>
<td>−0.984</td>
<td>0.748</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>0.518</td>
<td>0.967</td>
<td>−0.656</td>
<td>−0.907</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ</td>
<td>0.505</td>
<td>1.000</td>
<td>−0.726</td>
<td>−0.984</td>
<td>0.967</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>UE/E</td>
<td>−0.346</td>
<td>−0.873</td>
<td>0.316</td>
<td>0.858</td>
<td>−0.846</td>
<td>−0.873</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: ALP, f, δ, u, θ, UE/E and p₁ 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, 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.
M Business Cycle Statistics in the Standard Model

We provide further details on the performance of the standard model, which does not feature countercyclical congestion, both with and without separation shocks. For convenience, we repeat here business cycle statistics in the data (Panel A of Table A7). In the main text, Table 2 documents that the congestion model succeeds in replicating these patterns. In contrast, Table A7 documents the failures of the standard model.

In particular, the standard DMP model without separation shocks cannot match the cyclicality of UE flows, as explained in Section 2.1. Instead of a strong positive correlation with unemployment (0.74), the standard model without separation shocks predicts a correlation −0.27. In addition, as is well known, the model fails dramatically in replicating the volatility of labor market variables (Shimer, 2005). For instance, the volatility of labor market tightness in the model is only 7 percent of that found in the data.

Incorporating separation shocks into the standard model helps along several dimensions. Most notably, the correlation of UE flows and unemployment becomes positive and close to that in the data (0.74). In addition, because of the extra fluctuations in separations, other labor market variables become more volatile, but still fall short of their empirical volatilities. For instance, the standard deviation of labor market tightness becomes almost 20 percent of that in the data. However, these improvements come at a cost: the Beveridge curve turns counterfactually positive. The correlation coefficient between unemployment and vacancies in the model with separation shocks is 0.96, instead of the strongly negative (−0.934) correlation found in the data.\footnote{In the standard model (with and without separation shocks), average labor productivity is equal to the marginal productivity of new hires, as there is no distinction between worker types, so we omit this entry.}

In Appendix I, we analytically solve for the elasticity of labor market tightness to the separation rate, and show that this elasticity is small in a broad class of model parameterizations, echoing the results we present in this appendix.
Table A7: Business Cycle Properties of the No-Congestion Model

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Standard model without separation shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.004</td>
<td>0</td>
<td>0.003</td>
<td>0.013</td>
<td>0.015</td>
<td>0.003</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.704</td>
<td>0.704</td>
<td>0</td>
<td>0.843</td>
<td>0.592</td>
<td>0.704</td>
<td>0.306</td>
</tr>
</tbody>
</table>

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>1.000</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>-0.643</td>
<td>-0.643</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>0.980</td>
<td>0.980</td>
<td></td>
<td>-0.481</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>1.000</td>
<td>1.000</td>
<td>-0.643</td>
<td>0.980</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ</td>
<td>0.476</td>
<td>0.476</td>
<td>-0.272</td>
<td>0.476</td>
<td>0.476</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>UE/E</td>
<td>-0.348</td>
<td>-0.402</td>
<td>0.413</td>
<td>0.739</td>
<td>0.747</td>
<td>-0.402</td>
<td>1</td>
</tr>
</tbody>
</table>

**Panel B: Standard model with separation shocks**

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.010</td>
<td>0.005</td>
<td>0.088</td>
<td>0.068</td>
<td>0.058</td>
<td>0.017</td>
<td>0.067</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>0.688</td>
<td>0.647</td>
<td>0.499</td>
<td>0.736</td>
<td>0.751</td>
<td>0.647</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>ALP</th>
<th>f</th>
<th>δ</th>
<th>u</th>
<th>v</th>
<th>θ</th>
<th>UE/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>0.975</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>-0.441</td>
<td>-0.627</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>-0.508</td>
<td>-0.665</td>
<td>0.916</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>-0.306</td>
<td>-0.482</td>
<td>0.888</td>
<td>0.974</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ</td>
<td>0.975</td>
<td>1.000</td>
<td>-0.627</td>
<td>-0.665</td>
<td>-0.482</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>UE/E</td>
<td>-0.348</td>
<td>-0.402</td>
<td>0.413</td>
<td>0.739</td>
<td>0.747</td>
<td>-0.402</td>
<td>1</td>
</tr>
</tbody>
</table>

**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. Top panel reports values from the data, the bottom two panels from the standard model. All variables have been logged and detrended using the HP-filter with a smoothing parameter of 1,600.
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 A17.

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 $u^z$,

$$u^z_{t+1} = (1 - f(\theta^z_t))u^z_t + \delta(1 - u^z_t), \quad \kappa = q(\theta^z_t)\beta \mathbb{E}(1 - \delta)S^z_{i,t}$$

and, respectively, for $u^δ$,

$$u^δ_{t+1} = (1 - f(\theta^δ_t))u^δ_t + \delta_t(1 - u^δ_t), \quad \kappa = q(\theta^δ_t)\beta \mathbb{E}(1 - \delta)S^δ_{i,t+1}$$

Figure A18 plots the associated time series of these counterfactual unemployment rates together with actual unemployment. Table A8 provides a set of business cycle statistics related to overall unemployment and the three counterfactuals.

**Volatility.** Table A8 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 percent of the historical unemployment rate fluctuations in the United States. Its standard deviation is around 40 percent of the empirical one.⁹

**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 4.6.

⁹As discussed in Section 4.3, our model matches UE flows by estimating a somewhat more volatile separation rate process. In Table A8, this property leads the model to exaggerate 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 A17: Time Paths of Labor Market Variables

(a) ALP
(b) UE Share
(c) Job Separation Rate
(d) Job Finding Rate
(e) Unemployment
(f) Vacancies

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 A18: Unemployment Components

(a) Separations
(b) TFP Fluctuations

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 (A21) and (A22).

Table A8: Historical Decomposition of Unemployment: Model and Counterfactuals

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Congestion only</th>
<th>z only</th>
<th>δ only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.124</td>
<td>0.050</td>
<td>0.004</td>
<td>0.088</td>
</tr>
<tr>
<td>Contribution to total</td>
<td>1</td>
<td>0.297</td>
<td>0.008</td>
<td>0.657</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.905</td>
<td>0.950</td>
<td>0.865</td>
<td>0.825</td>
</tr>
<tr>
<td>corr($x, y$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congestion only</td>
<td></td>
<td>0.729</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>z only</td>
<td></td>
<td>0.274</td>
<td>−0.264</td>
<td>1</td>
</tr>
<tr>
<td>δ only</td>
<td></td>
<td>0.920</td>
<td>0.411</td>
<td>0.464</td>
</tr>
</tbody>
</table>

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.