

A Congestion Theory of Unemployment Fluctuations

Yusuf Mercan, Benjamin Schoefer, Petr Sedláček

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

A Congestion Theory of Unemployment Fluctuations

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, diminishes their marginal product and discourages further job creation. Countercyclical congestion alone explains about 30–40 percent of U.S. unemployment fluctuations. Besides generating realistic labor market volatility, it also provides a unified explanation for the cyclical labor wedge, the excess earnings losses from job displacement and from graduating during recessions, and the insensitivity of unemployment to labor market policies, such as unemployment insurance.

JEL-Codes: E240, J630, J640.

Keywords: unemployment, business cycles, recessions.

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

Benjamin Schoefer
University of California Berkeley / CA / USA
schoefer@berkeley.edu

Petr Sedláček
University of Oxford / United Kingdom
petr.sedlacek@economics.ox.ac.uk

November 22, 2020

We thank Chris Edmond, Michael Elsby, Domenico Ferraro, Shigeru Fujita, Bart Hobijn, Philip Jung, Fatih Karahan, Simon Mongey, Emi Nakamura, Aysegül Sahin, 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, and the University of Adelaide. 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.

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.² 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 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. Similar patterns can be found in other OECD countries. Yet, while countercyclical unemployment to employment flows are a robust empirical fact (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

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

²The notion that the relative supply of different labor inputs is relevant for long-run macroeconomic outcomes has a long tradition (see e.g. Katz and Murphy, 1992; Krusell, Ohanian, Rios-Rull, and Violante, 2000; Card and Lemieux, 2000; Jeong, Kim, and Manovskii, 2015). Our paper proposes that a similar mechanism is important also for business cycle analysis.

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 finding in the aggregate time series is 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). This property stands in sharp contrast to the standard search models, which exhibit no congestion, making firms quickly hire away such an increase in job losers.³

Our congestion theory of unemployment integrates both these empirical facts into an otherwise standard Diamond-Mortensen-Pissarides (DMP) search and matching model of the labor market. The combination of these two features delivers countercyclical congestion. First, shocks to separations generate countercyclical UE flows. Second, our aggregate production function features diminishing returns in the size of a given cohort of new hires, i.e., congestion. Together, these two features rationalize why in a recession, firms do not hire away the additional job losers.

We formalize diminishing returns in hiring by assuming that different cohorts of workers are not perfect substitutes for one another. For example, different cohorts may be on different rungs of the job ladder, have different experience or firm-specific skill levels, and hence perform different tasks.⁴ 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 of Shimer (2005). We discipline this parameter by having our model match the empirical impulse response of hiring (labor market tightness) to a separation rate shock.⁵ In the standard, no-congestion model, this response is counterfactually flat.

³In the standard Diamond-Mortensen-Pissarides (DMP) model, hiring (vacancies) simply scales one to one with any change in unemployment that leaves other fundamentals—such as productivity—unchanged. Concretely, the response of labor market tightness to separations is flat.

⁴Going back to Doeringer and Piore (1985), there is also an empirical and theoretical literature in organizational economics that has emphasized job structures and production functions that render entry-level jobs (“ports of entry”) imperfect substitutes for higher-tier jobs and job-specific diminishing returns (for firm-level evidence, see, e.g., Doran, Gelber, and Isen, 2020; Jäger and Heining, 2019). Our model, which focuses on movements between unemployment and employment, considers job-to-job transitions to leave job switchers on track on the job ladder and does not entail skill loss.

⁵Here, our calibration strategy echoes the important prior work by Coles and Moghaddasi Kelishomi (2018), who relax the free-entry condition in accordance with the data and show that separation shocks

Studying the impulse response of labor market tightness to separation shocks therefore provides a clear target pinning down the empirical degree of congestion.

Importantly, any modeling route that results in congestion would inherit the amplification properties of our model. We show this robustness property by means of “iso-congestion” models, which have different underlying mechanisms of congestion but crucially each are calibrated to match the empirical degree of congestion (the impulse response of labor market tightness to separation rate shocks).⁶ We therefore view our theory as representing a larger class of congestion models. At the same time, however, our baseline model is also consistent with a range of other macroeconomic patterns, which we discuss below, providing additional external validity of our specific modelling choice.

Estimating our calibrated model on US time series data, we find that countercyclical congestion alone accounts for more than 30–40 percent of observed unemployment fluctuations, and much of its persistence. The full model, which also features standard total factor productivity (TFP) 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 in our full model 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. We show that this additional volatility can easily be masked in the standard measure of average labor productivity. Indeed, our framework is consistent with a modestly volatile average labor productivity time series, while matching the highly volatile labor market variables—a long-standing macroeconomic challenge (see, e.g., Shimer, 2005; Hagedorn and Manovskii, 2008; Pissarides, 2009; Hall, 2017; Ljungqvist and Sargent, 2017).

The second feature key to the quantitative success of our model is the presence of cohort effects, through which aggregate conditions at the time of hiring have long-lasting effects on a given cohort’s productivity. Consider a cohort of unemployed hired during a recession. Its relative abundance in employment diminishes the marginal products of its

affect unemployment.

⁶For example, alternatives with convex adjustment costs in total hiring (see e.g., Fujita and Ramey, 2007; Coles and Moghaddasi Kelishomi, 2018; Mercan and Schoefer, 2020) exhibit similar cyclical properties. Similarly, specifications with different skill evolution during unemployment (which have been explored in models *without congestion* by, e.g., Ljungqvist and Sargent, 1998, 2004; den Haan, Haefke, and Ramey, 2005) would also deliver the same model behavior, once the congestion parameter in the production function is recalibrated to match the empirical degree of congestion.

members. Until they become unemployed again, these workers never escape their cohort's abundance and diminished productivity, even after the economy has long recovered.

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).⁷ 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. In fact, our framework nests the model in Shimer (2005) as a special case with perfect substitution between cohorts of workers.

In addition to a new perspective on unemployment fluctuations, our framework offers solutions to three long-standing macroeconomic challenges linked to unemployment fluctuations. These results provide further external validity for our modeling choices.

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; Bilts, Klenow, and Malin, 2018). As a procyclical multiplier on the standard MPL measure, the amplification of new hires' productivity fluctuations manifests itself precisely as a countercyclical labor wedge between the MRS and the standard MPL measure.

Second, our model features large, countercyclical, and persistent earnings losses from job displacement and from labor market entry such as from university graduation. In our model, consistent with the data (Davis and von Wachter, 2011), job separations during recessions come with earnings losses of about 15 percentage points more severe than those in booms, which do not fully fade even ten years after the event. Similarly, our model also generates realistic scarring effects of graduating in recessions (Kahn, 2014; Oreopoulos, von Wachter, and Heisz, 2012; Schwandt and von Wachter, 2019). Both these model predictions rest on the presence of strong cohort effects, which yields independent validation of our parameterization. Moreover, these earnings losses largely stem from persistent wage declines and "lower quality" jobs (Oreopoulos, von Wachter, and Heisz, 2012; Schmieder, von Wachter, and Heining, 2019; von Wachter, forthcoming), consistent with the productivity channel in our baseline model.

⁷Michaillat (2012) presents a model with wage rigidity and diminishing returns in total employment, but homogeneous workers. While the focus of that paper is different—to show that much of unemployment is driven by rationing due to wage rigidity, rather than search frictions—the model does not exhibit congestion in that it would predict essentially no effect of separation rate shocks on labor market tightness, as in the standard DMP model.

Third, because our model obtains amplification through more volatile allocative productivity (rather than a high elasticity to productivity changes), we overcome the critique raised by Costain and Reiter (2008), that standard DMP models cannot simultaneously exhibit high volatility of unemployment driven by productivity shocks, and a low sensitivity to policy changes such as in unemployment insurance (UI) generosity. Our model features a long-run elasticity of unemployment with respect to UI within their estimated range—exactly because amplification in our framework does not rely on a low fundamental surplus (Hagedorn and Manovskii, 2008; Ljungqvist and Sargent, 2017).

Our paper further relates to several studies that analyze associated mechanisms and empirical patterns. Dupraz, Nakamura, and Steinsson (2019); Hall and Kudlyak (2020b,a) study the time series properties of the unemployment rate with the focus on the dynamics of the gradual reduction in unemployment during the recovery, properties that our congestion model improves on compared to a standard DMP model by curbing UE flows and generating persistence. A notion of congestion is present in important prior work by Coles and Moghaddasi Kelishomi (2018), who relax the free-entry condition—making vacancy creation less than infinitely elastic, and highlighting the role of separation shocks. Hall (2005a) and Engbom (2020) provide models in which the unemployed send applications in less selective ways in recessions, such that recruitment becomes more difficult and costly, a process that can be interpreted to reflect congestion. However, we are not aware of a prior comprehensive empirical and theoretical analysis of congestion in hiring, its role in unemployment fluctuations, and its effect on a range of other macroeconomic patterns. 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

We provide two pieces of empirical evidence that point to countercyclical congestion in new jobs. First, we document that during and in the aftermath of economic downturns, the composition of employed workers shifts toward those with recent unemployment spells. We trace the origins of this pattern to countercyclical unemployment-to-employment (UE) worker flows, which we show to be driven by countercyclical job separations. Second, we present time series evidence for the limited capacity of firms in absorbing these UE hires,

i.e., for congestion in hiring. We additionally discuss existing cross-sectional, local-labor-market and firm-level, evidence for congestion.

2.1 Countercyclical Unemployment-to-Employment Flows

The first ingredient generating countercyclical congestion is the countercyclical employment share of workers with recent unemployment spells. We trace this pattern back to countercyclical UE worker flows and job separation rates.

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 of workers with recent unemployment drawing on 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 countercyclicity by plotting the log deviation in our employment 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).

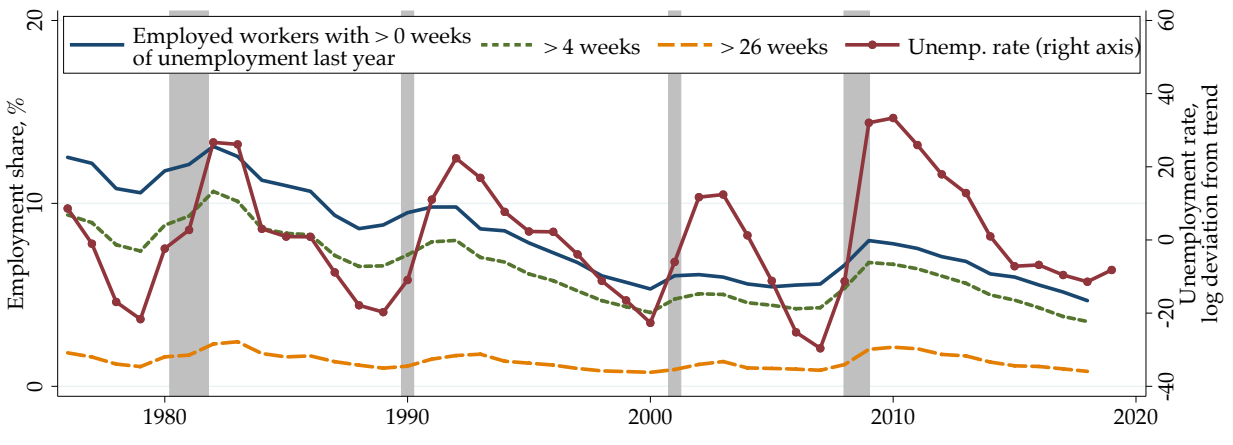
We quantify the countercyclicity as an Okun’s law: the elasticity of the new-hire share in employment with respect to the unemployment rate is 0.493. We visualize this relationship in the scatter plot in Panel (c) of Figure 1.

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).

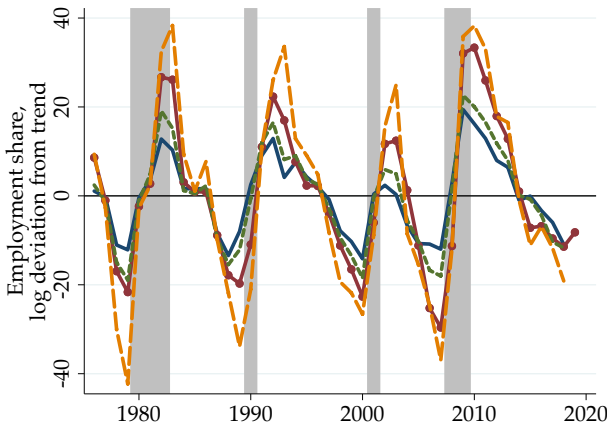
We start by documenting that monthly unemployment-to-employment (UE) worker flows are countercyclical in Figure 2 Panel (a). Here, we draw on monthly CPS data covering 1976m1–2019m12. We track individuals switching their labor force status from one month to the next using the rotating-panel structure of the CPS. We construct quarterly averages of the monthly transition rates and only for visual clarity smooth the time series

Figure 1: Countercyclicity of the Employment Share with Unemployment Past Year

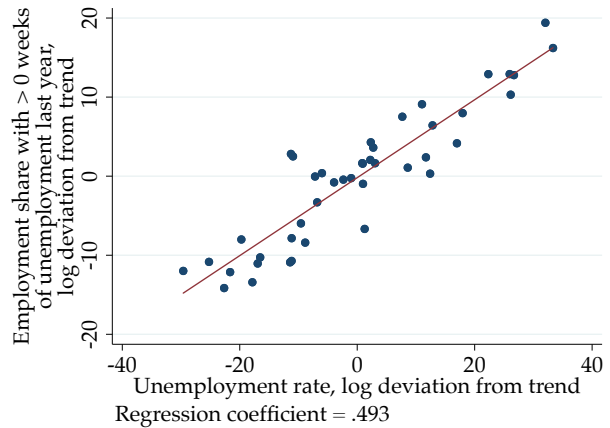
(a) Employment Shares of Workers with Unemployment Last Year by Total Weeks, and Unemployment Rate



(b) Cyclicity: Log Deviations from Trend



(c) Okun's Law

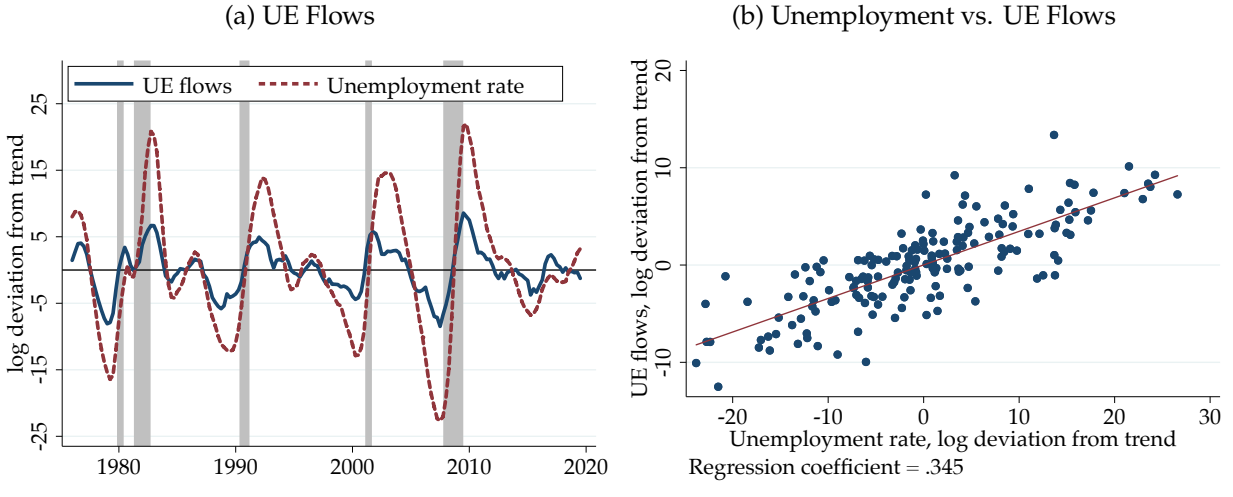


Notes: Panel (a) plots the share of employed workers who have undergone unemployment in the preceding calendar year for different amount of weeks (total). Panel (b) plots their log deviations from trend. Panel (c) reports the scatter plot of the detrended time series. The time series are HP filtered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

by taking four-quarter centered moving averages (but we use the underlying quarterly data for any statistic we report). We largely follow Fujita and Ramey (2006) in these steps, relegating further details about data construction, sample selection and measurement into Appendix A.

UE worker flows expand dramatically during all U.S. recessions since 1976, moving tightly with the unemployment rate. Panel (b) 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

Figure 2: The Countercyclicality of Unemployment-to-Employment (UE) Flows



Notes: Panel (a) plots the log deviations in UE flows and log deviations in the unemployment rate from their respective trends. Both series are based on quarterly averages of monthly data and for visual clarity are smoothed by taking centered four-quarter moving averages. Panel (b) plots log deviations in UE flows against log deviations in the unemployment rate. Detrended series are HP filtered with a smoothing parameter of 1600. Shaded regions denote NBER-dated recessions. Source: CPS monthly files.

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).

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, fraction δ (“separation rate”) of employed workers separate into unemployment, and fraction f (“job finding rate”) of unemployed job searchers find, and accept, a job.⁸ This bathtub model of “ins” and “outs”

⁸In the data, and later on in the model, we specify discrete-time transition probabilities while using the conventional term “rates” interchangeably.

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:

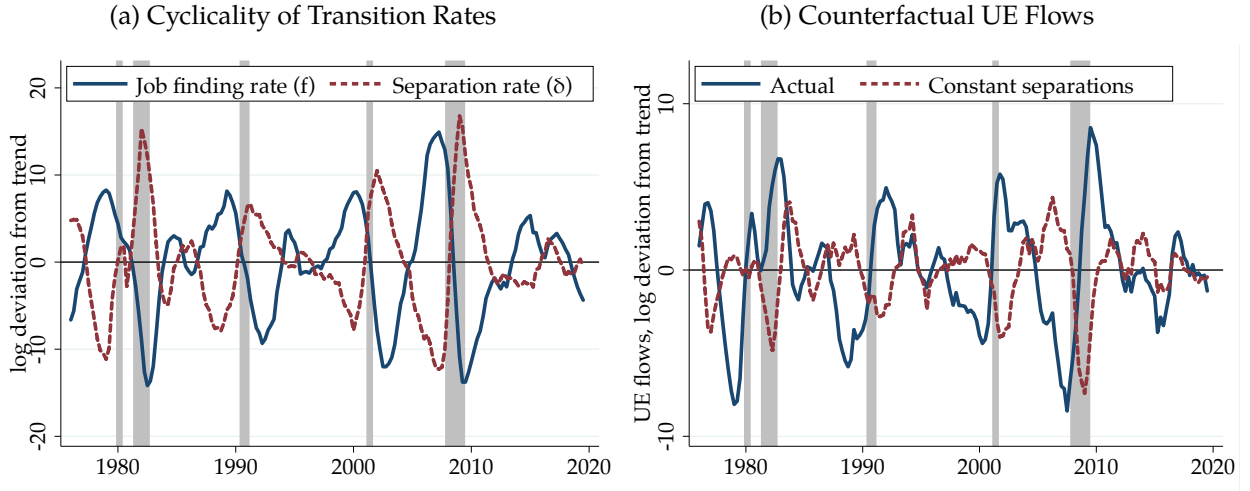
$$\frac{dUE/UE}{du/u} = \frac{df/f}{du/u} + 1 = \frac{1}{(1-u) \left[-1 + \frac{d\delta/\delta}{df/f} \right]} + 1, \quad (4)$$

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 U.S., 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 $-u/(1-u) \approx -0.067$ for an average US unemployment rate of $u \approx 0.063$.

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

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 1600 and smoothed by taking centered four-quarter moving averages for visual clarity. Shaded regions denote NBER-dated recessions. Source: CPS monthly files.

law of motion for unemployment, using the observed job finding rate yet holding the separation rate at its sample average $\bar{\delta}$.⁹ 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 $\frac{dUE}{UE} = u \frac{df}{f} + (1-u) \frac{d\delta}{\delta}$. Here, in percent terms, movements in the separation rate δ 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 and Within-Period EUE Flows. For consistency with the discrete time model that we present below, the empirical transition rates are not adjusted for time aggregation. Appendix C shows robustness from using time-aggregation-adjusted transition rates to impute our main time series for UE flows. That time series includes additional transitions from initially employed workers who separate into unemployment and transition back into employment within the period. That is, this method

⁹Specifically, we iterate on the law of motion for unemployment given by $\tilde{U}_{t+1} = (1 - f_t)\tilde{U}_t + \bar{\delta}(L_t - \tilde{U}_t)$ to construct the counterfactual unemployment time series \tilde{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 $\tilde{UE}_t = f_{t-1}\tilde{U}_{t-1}$.

replicates (at the monthly frequency) the CPS ASEC definition of asking the end-of-period employed about potential unemployment spells during the period. We find very similar results for the cyclical behavior of these UE flows adjusted for time aggregation, implying that our discrete-time definition is robust to this extension.

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), and therefore total (rather than those only out of unemployment) hires are not countercyclical. 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. Therefore, we sidestep job-to-job transitions in our empirical analysis, which are more likely to leave workers on track on the job ladder, whose specific human capital are less likely to be lost, and who likely enter non-entry level positions in contrast to new hires out of unemployment.

Unemployment vs. Nonemployment. Our framework is a standard two-state search model featuring unemployment and employment. The labor market in reality features a third state, namely workers out of the labor force, with flows to and from unemployment and employment (for a rich cyclical model of these flows, see Krusell, Mukoyama, Rogerson, and Şahin, 2017). As a robustness check, 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 countercyclicity 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 below in Section 2.2, with the model calibration strategy described in Section 4.2.

2.2 Evidence for Congestion Effects in Hiring

Having documented the countercyclical nature of the employment share of recently unemployed hires and of UE flows, 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.

Defining Congestion. We define our congestion concept as the economy’s limited capacity to absorb “pure disturbances” in the unemployment pool (i.e., that leave fundamentals

such as productivity and the discount factor constant) by means of adjusting hiring out of unemployment (i.e., UE flows) in the short run. While perhaps an intuitive property, it turns out that the standard DMP model—the canonical macroeconomic framework to analyze worker flows and unemployment—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 the short-run spikes in unemployment by increasing UE hires.

In both models, labor market tightness $\theta = v/u$, the ratio of vacancies v to the unemployed u , determines the job finding rate $f(\theta)$, with $f'(\theta) > 0$, $f''(\theta) < 0$, as matches are governed by a constant-returns-to-scale matching function $M(u, v)$.¹⁰ Unemployment evolves according to the following law of motion:

$$u_{t+1} = (1 - f(\theta_t)) u_t + \delta_{t+1}(1 - u_t). \quad (5)$$

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 scale one to one with unemployment, such that their ratio, labor market tightness θ , is fixed. The reason is constant returns in production as well as in the matching function. Here, θ moves around only in response to shifts in factors that affect either the benefit (e.g., productivity net of wages, the discount factor) or the costs of hiring.

An important implication of the absence of congestion in hiring is that pure shifts in the amount of the unemployed have no effect on θ and hence on the individual-level job finding rate. Since adjustment is fast in the model, the unemployment rate quickly converges back to steady state, such that the newly unemployed are absorbed into the economy nearly immediately. In fact, for this economy, we can provide analytical adjustment paths following a perfectly transitory separation shock hitting at $t = 0$:

$$\begin{aligned} u_{t+1} &= u_t - f(\theta_{ss})u_t + (1 - \delta_{ss})(1 - u_t), \\ v_t &= \theta_{ss} \cdot u_t, \quad \theta_t = \theta_{ss}, \quad UE_t = u_t \cdot f(\theta_{ss}). \end{aligned} \quad (6)$$

The half life for unemployment to recover, $t_{0.5}^{nc}$, i.e., the time it takes to arrive at $\frac{u_{t_{0.5}} - u_{ss}}{u_0 - u_{ss}} = 0.5$,

¹⁰Foreshadowing our quantitative exercises, we adopt the standard Cobb-Douglas matching function, $M(u, v) = \bar{m} u^{1-\mu} v^\mu$ (for which $f(\theta) = \mu \theta^{1-\mu}$, where \bar{m} denotes matching efficiency and μ is the matching elasticity). We set $\bar{m} = 0.57$ and $\mu = 0.72$ based on our preferred model calibration.

in the no-congestion model is

$$u_t - u_{ss} = (1 - (f_{ss} + \delta_{ss}))^t \cdot (u_0 - u_{ss}) \quad (7)$$

$$\Leftrightarrow \frac{u_t - u_{ss}}{u_0 - u_{ss}} = (1 - (f_{ss} + \delta_{ss}))^t \quad (8)$$

$$\Rightarrow t_{0.5}^{nc} = \log(0.5) / \log(1 - (f_{ss} + \delta_{ss})). \quad (9)$$

Since labor market fluidity is high in the United States, with $f \approx 0.570$ and $\delta \approx 0.042$ per quarter on average, this half life is small, around 0.731 quarters.

There are three takeaways from the non-congestion model. First, vacancies and unemployment move in the same direction (to hold θ and the job finding rate constant) in response to such separation shocks. Second, a no-congestion economy nearly immediately and completely absorbs disturbances in unemployment, such as those arising from separation rate shocks, keeping f unaffected. Third, the means by which the economy absorbs these shocks is a spike in UE flows, the shape of which inherits the shape of unemployment (due to constant f).

We illustrate these 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, which keeps labor market tightness constant. Hence, the job finding rate is constant, leading UE flows to spike, which is exactly the mechanism that achieves quick 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 such as that 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, which makes immediately clear why labor market tightness and the job finding rate must *fall* when unemployment inflows increase due to congestion:

$$\begin{aligned} u_{t+1} &= u_t - f(\theta_t)u_t + (1 - \delta_{ss})(1 - u_t), \\ v_t &= \theta_t \cdot u_t, \quad \theta_t = f^{-1}(UE_{ss}/u_t), \quad UE_t = f(\theta_t)u_t = UE_{ss}. \end{aligned} \quad (10)$$

Following similar steps to the no-congestion case, we derive the half life of unemployment recovery in the full-congestion model, $t_{0.5}^{fc}$, as follows:

$$u_t - u_{ss} = (1 - \delta_{ss})^t \cdot (u_0 - u_{ss}) \quad (11)$$

$$\Leftrightarrow \frac{u_t - u_{ss}}{u_0 - u_{ss}} = (1 - \delta_{ss})^t \quad (12)$$

$$\Rightarrow t_{0.5}^{fc} = \log(0.5)/\log(1 - \delta_{ss}). \quad (13)$$

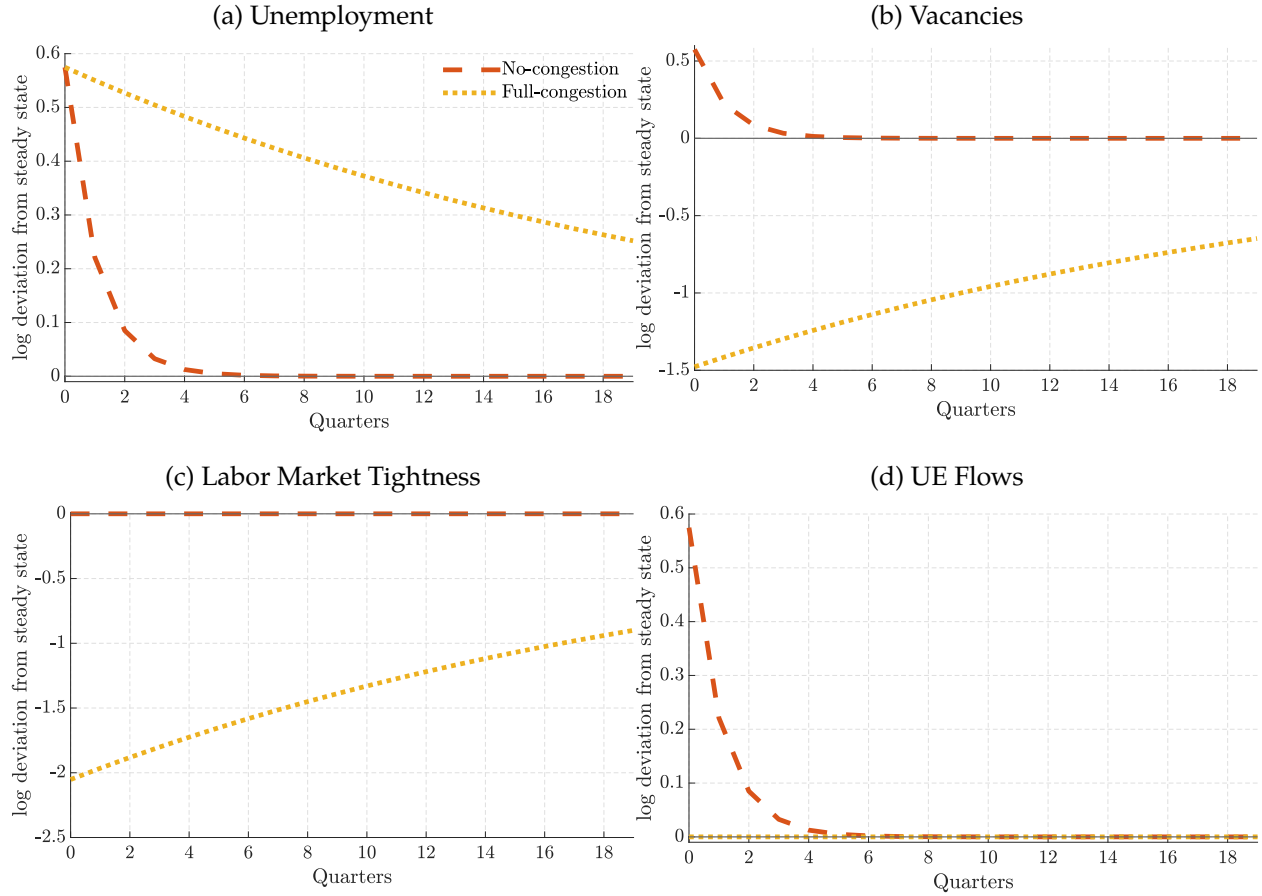
Calibrated to the US average $\delta = 0.0425$, this half life is around 16 quarters. That is, the recovery of unemployment is around 20 times faster in the no-congestion model compared to the full-congestion one. In fact, the full-congestion economy requires around 20 times the time periods the no-congestion economy does for an initial disturbance to decay to *any* fraction d (where setting $d = 0.5$ gives the half life):

$$t_d^{fc}/t_d^{nc} = \frac{\log(d)/\log(1 - \delta_{ss})}{\log(d)/\log(1 - \delta_{ss} - f_{ss})} \approx \frac{\delta_{ss} + f_{ss}}{\delta_{ss}} = \frac{1}{u_{ss}}. \quad (14)$$

Figure 4 plots, with the dotted yellow lines, the transition paths of this second extreme case with full congestion following an expansion in the unemployment pool after a perfectly transitory job separation shock. While, upon impact, unemployment increases by the same amount as in the no-congestion benchmark, the transition dynamics differ dramatically. UE flows are by the nature of this full-congestion economy constant, rather than increasing sharply as in the no-congestion economy. To achieve constant gross hiring (UE flows), the matching function requires vacancies to *fall*, since hiring the same amount of workers is easier due to the abundance of the unemployed. Hence, each individual unemployed worker's job finding rate falls, as the same amount of hires are spread across more unemployed, or, in matching function terms, labor market tightness falls (vacancies fall while unemployment increases). Consequently, unemployment recovers only very slowly, remaining high even many months after the transitory separation shock.

Aggregate Time Series Evidence. Using a vector autoregression (VAR) model, we show that the aggregate behavior of labor market variables rejects the prediction of no-congestion models. To diagnose congestion in the data, we primarily focus on labor market tightness exactly because it mediates hiring in the DMP model. It is also the variable in which no-congestion and congestion models most starkly differ in their responses to (EU) separation shocks, as shown above. Our empirical exercise follows the thought experiment presented in Figure 4, and primarily studies the response of labor market tightness

Figure 4: Congestion in Hiring: Impulse Responses to a Transitory Separation Shock



Notes: The figure plots the impulse responses of unemployment, vacancies, labor market tightness and UE flows to a one-percent, perfectly transitory job separation shock in economies that feature no- and full congestion in hiring.

to a separation-shock induced expansion in the unemployment pool in the aggregate time series.

Specifically, we study the behavior of two sets of endogenous variables given by the vector:

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

where ALP is average labor productivity (measured as output per worker in the non-farm business sector), δ is the separation rate (EU flows divided by beginning-of-period employment), and 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 \mathbf{y}_t :

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}(L)\mathbf{y}_{t-1} + \mathbf{v}_t, \quad (15)$$

where \mathbf{c} is a constant term, $\mathbf{A}(L)$ is a lag polynomial, and $\mathbf{v}_t \sim (0, \mathbf{\Omega})$ is a vector of error terms with variance-covariance matrix $\mathbf{\Omega}$. We include four lags of the endogenous variables in our specification. We identify productivity and separation shocks using a recursive identification scheme (or, equivalently, using a Cholesky decomposition of $\mathbf{\Omega}$). Our timing assumptions are that *ALP* has a contemporaneous effect on both δ and x , and that δ only has a contemporaneous effect on x , whereas x affects the endogenous variables with a lag. We then study impulse responses to an orthogonalized shock to δ , to isolate the effect of movements in job separations from that of productivity fluctuations.¹¹

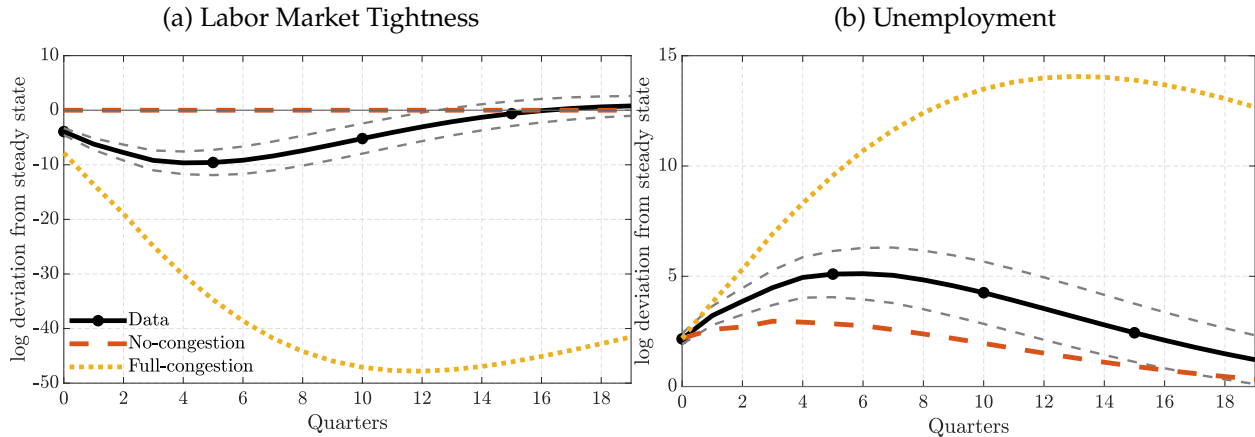
Figure 5 plots the empirical impulse response functions of labor market tightness (Panel (a)) and unemployment (Panel (b)) to a separation shock. We also report the two counterfactual benchmark responses from an economy with no congestion (red dashed line), and full congestion in hiring (yellow dotted line), following the definitions laid out above. These two benchmarks mimic those in Figure 4 but they use the estimated IRFs of labor market tightness and are in response to non-transitory, empirical separation rate shocks.

Again, the no-congestion benchmark implies a fixed labor market tightness because the separation rate shock does not shift fundamentals (productivity) and so vacancies scale one to one with unemployment; as a result, unemployment rises only little and recovers quickly.

The data clearly reject the insensitivity of labor market tightness that characterizes the no-congestion—and hence DMP—benchmark. The empirical response of labor market tightness is significantly negative and persistent (Panel (a)). That is, when the unemployment pool expands (such as due to EU separations that leave standard fundamentals such as productivity constant), the economy does not expand job opportunities to absorb the newly unemployed workers nearly immediately, but responds only with a small increase

¹¹This orthogonality with productivity holds exactly in the first period. In Appendix Figure A7, we present the IRFs of *ALP* to the δ shock. Importantly, if anything, the empirical process indicates (insignificantly) positive *ALP* responses to a positive separation rate shock in the transition periods. Hence, the comovement of productivity with the separation shock would lead to an increase rather than decrease of labor market tightness (and a decrease in unemployment). Moreover, evidence suggests that the composition of the unemployment pool improves and that firms find it profitable to increase their hiring standards in recessions (see e.g. Mueller, 2017; Modestino, Shoag, and Ballance, 2016). Congestion arises in our model as long as the pool of the unemployed differ from the employed. See den Haan et al. (2000) for a treatment of turbulence with endogenous separations, and Ferraro (2018) for a model with permanent worker heterogeneity in productivity and endogenous separations.

Figure 5: Congestion: Empirical Impulse Responses to a Separation Shock



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.

in hiring. Instead of a constant job finding rate, individual unemployed workers find it dramatically harder to find jobs. The resulting drop in job finding rates, paired with the increase in separations, triggers a larger and more persistent increase in unemployment (Panel (b)). These empirical patterns are absent in the language of the standard DMP model.

Of course, quantitatively, the empirical responses of labor market variables still lie *in between* the two extremes of the no-congestion and full-congestion benchmarks. Specifically, while the drop in labor market tightness is qualitatively consistent with the full-congestion view, the empirical responses of labor market tightness and unemployment do recover after some quarters, in contrast to the extreme persistence in the full-congestion benchmark. In our quantitative model presented in Section 3, we therefore 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).

Related Quasi-Experimental Evidence. Besides aggregate time series evidence, firm-level and local-labor market evidence are consistent with congestion. At the firm level, Doran, Gelber, and Isen (2020) draw on quasi-experimental variation in recent hires arising from U.S. visa lotteries, and find that an exogenously assigned new hire (more than) fully crowds out any additional subsequent hiring into that job type, which would imply full congestion at the firm level.

Zooming into local labor market adjustment, Mercan and Schoefer (2020) review 15 studies and document very limited short-run employment spillovers from particular firms in a local labor market onto peer firms unaffected by the first group’s labor demand shifters. For example, policy incentives targeting some eligible firms do not affect hiring by ineligible employers in the same local labor market (Cahuc, Carcillo, and Le Barbanchon, 2018); similarly, sharp labor demand reductions and mass layoffs by particular plants or sectors do not lead other employers to expand (e.g., Mian and Sufi, 2014; Gathmann, Helm, and Schönberg, 2018). That is, extensive margin adjustment (including EU separation events) do not appear to spur hiring in unaffected firms, consistent with considerable degree of congestion.

3 A Search Model with Countercyclical Congestion

We now integrate countercyclical congestion into an otherwise standard DMP model. In Sections 4 and 5, we study our calibrated model quantitatively and explore its success in rationalizing a number of cyclical labor market facts.

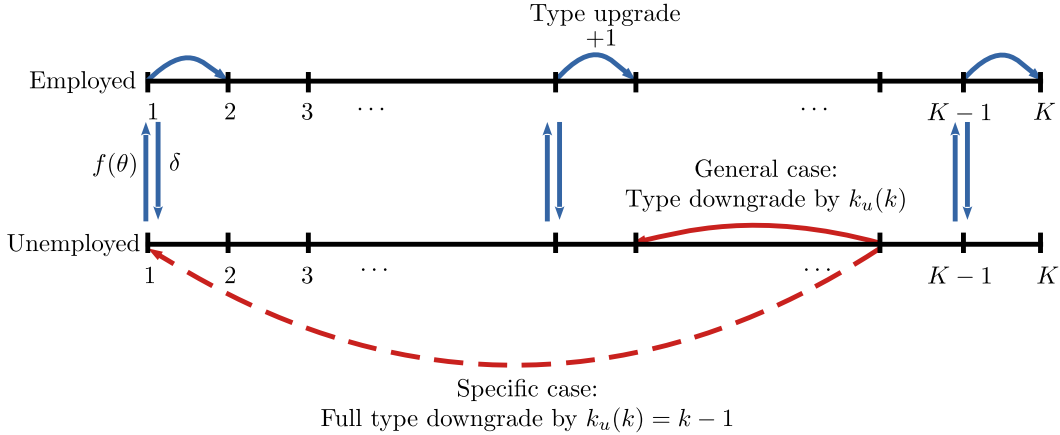
To generate congestion in jobs that UE hires fill, the model features only two additional ingredients into the canonical DMP framework. First, to generate countercyclical UE flows, which underlie our first empirical finding of countercyclical UE hires share in employment, the model features countercyclical separations, consistent with evidence presented in Section 2.1. Second, to obtain congestion dynamics as suggested by the evidence in Section 2.2, our model features diminishing returns in new hires. We achieve this property by introducing imperfect substitution between hiring cohorts in an aggregate production function.¹² 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 as many laid off workers as quickly as predicted by no-congestion models. We also show robustness to alternative sources of congestion such as convex hiring costs.

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, which generate the diminishing returns in hiring as the source of congestion in our main setup.

¹²This modelling choice is akin to the seminal analyses of long-run labor market changes by Katz and Murphy (1992) and Card and Lemieux (2000) .

Figure 6: Worker Type Evolution



Worker Types. Workers are heterogeneous in their type $k \in \mathcal{K} = \{1, \dots, K\}$, 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. Examples of such mechanisms include skills gained on the job and job ladders.

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 $k_u(k) \in \{0, 1, \dots, k - 1\}$ 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 $k_u(k) = k - 1$.

Congestion: Production with Diminishing Returns to Worker Types. Worker heterogeneity matters through the production function. “Productivity” p_k is formally the price of the intermediate input, which is differentiated by worker type k . Intermediate goods $\{n_k\}_{k=1}^K$ 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

$$Y = z \left(\sum_{k=1}^K \alpha_k n_k^\sigma \right)^{1/\sigma}, \quad (16)$$

where α_k is a type-specific productivity shifter associated with type k , n_k is the stock of type- k workers in production, who operate a linear technology that converts one unit

of labor to a unit of the intermediate good of type k , and σ governs the elasticity of substitution between inputs, $\frac{1}{1-\sigma}$. This functional form exhibits overall constant returns to scale and a constant elasticity of substitution across worker types. In Appendix D, 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.

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

$$p_k = \alpha_k n_k^{\sigma-1} \frac{Y}{\sum_{l=1}^K \alpha_l n_l^\sigma} = \alpha_k s_k^{\sigma-1} \frac{1}{\sum_{l=1}^K \alpha_l s_l^\sigma} \frac{Y}{N}, \quad (17)$$

where $N = \sum_{l=1}^K n_l$ denotes aggregate employment, Y/N is the standard average labor productivity (ALP), and $s_l = n_l/N$ denotes the employment share of type- l workers. Equation (17) 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 worker-type ladder as one cohort. This case permits an easy representation of new hires' marginal product of labor, namely $p_{k=1}$, and its present value over long-term jobs.

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 downgrade. We plot this relationship for two levels of congestion parameter $\sigma \in \{0.241, 1\}$. We normalize steady-state marginal products to one for all worker types, i.e., $p_k = 1$ for all k , by means of adjusting α_k , so that worker heterogeneity purely matters through diminishing returns/congestion rather than through mechanical composition effects (e.g., Mueller, 2017), which we intentionally sidestep.

The flat yellow dotted line captures the case of $\sigma = 1$, for which workers are perfect substitutes, and each type's marginal product simply equals the average labor productivity, Y/N . Shifts in the share of new hires have no effect on labor productivity of either the new hires or the average worker. 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 $1 - 0.241 = 0.759$).

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 our choice of a CRS-CES production function where there are diminishing returns to specific worker types but the constant returns to scale over all worker types is preserved. This property is crucial to the empirical potential of our mechanism: the large fluctuations in productivity of new hires our model imply can be masked by—and hence be consistent with—the standard, relatively smooth ALP in the data.¹³

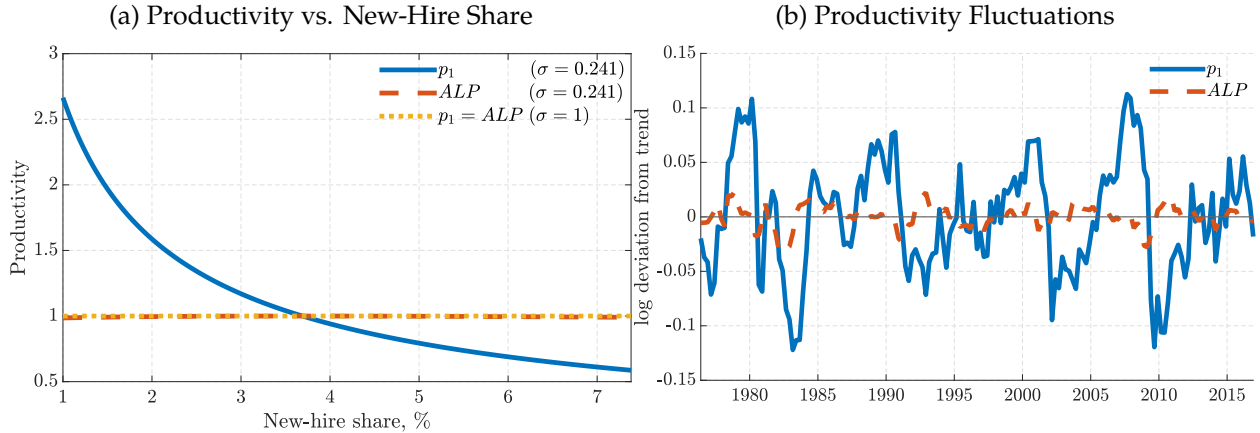
While specifying the full model, we present the general case regarding type downgrading, and then calibrate our model under the specific assumption above. We also show robustness to alternative type-downgrading specifications in Section 4.3; once recalibrated to match the same targets, these variants are *isomorphic*.

Preview: Countercyclical Congestion. Figure 7 Panel (b) illustrates the amplification from countercyclical congestion for this leading case of full downgrading to $k = 1$ for unemployed job seekers. It plots the time series (log deviations from trend) of productivity of new hires p_1 , along with the average labor productivity Y/N . We construct new-hire productivity p_1 by feeding in the observed share of UE hires, $s_{1,t}$, at each quarter, which gives $p_{1,t} = \alpha_1 s_{1,t}^{\sigma-1} ALP_t \frac{1}{\sum_{k=1}^K \alpha_k s_{k,t}}$, where ALP_t is the observed average labor productivity.¹⁴ 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).

¹³By calibrating all types' steady state productivity to equal one here (and in our full calibration), we abstract from compositional effects on productivity if, for example, new hires are less productive than incumbent workers. Such mechanical compositional effects would be present even with $\sigma = 1$.

¹⁴For 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



Notes: Panel (a) plots the marginal product of new hires and average labor productivity as a function of the employment share of new hires for different values of congestion parameter σ . Steady-state average labor productivity and each type’s marginal product are normalized to one for both calibrations of σ . Panel (b) plots the empirical US time series for average productivity and new-hire productivity. Both time series are in logs and detrended using an HP-filter with a smoothing parameter of 1600.

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 $\beta \in (0, 1)$. There are two types of producers: intermediate-input producers (which we conveniently call “firms” going forward as they feature prominently below), which use labor to produce output they sell in a perfectly competitive market to a final good producer (which we call “retailer” as it comes in solely to pin down intermediate input prices that stand for the marginal products of worker types). The retailer bundles the intermediate goods into a final consumption good using the technology in Equation (16) with total factor productivity (TFP) z . Individuals own both producer 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 $q(\theta) = \frac{M}{v} = M(1/\theta, 1)$.

No Job-to-Job Transitions. While we refer to the k -types as denoting either skill gained on the job or a rung on the job ladder, our model does not feature explicit job-to-job transitions. We sidestep this margin 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, whereas types reset upon separation into unemployment as in job ladder models or models of turbulence (Ljungqvist and Sargent, 1998, 2004; den Haan et al., 2005). 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).

Separations. Each period, active matches separate with exogenous but time-varying rate δ . These separations are an ad-hoc event rather than arising from endogenous decisions between the worker and firm in response to shocks to surplus. While we do not provide this extension, modeling endogenous separations should leave our key results intact provided such an extended model generates realistic separation and hiring cyclicalities, as well as, crucially, matching the impulse response of labor market tightness to EU separations unrelated to productivity movements (as documented in Figure 5).¹⁵

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

Timing. At the beginning of each period, aggregate productivity z and separation rate δ are realized. Worker-firm matches (both those active last period and those formed last period) are destroyed at rate δ , in which case the worker becomes unemployed and the firm becomes vacant. The surviving matches produce the intermediate inputs differentiated by the type of the worker k , which the retailer bundles into the final consumption good.

¹⁵For a discussion of endogenous separations in the context of turbulence models, see den Haan, Haefke, and Ramey (2005).

Workers consume their wage or unemployment benefits, depending on their employment status and k -type. Employed workers upgrade by one type, and unemployed workers downgrade by k_u types. The period closes by the search stage. Firms post vacancies and unemployed workers search for jobs, which determine market tightness. New matches are formed.

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

$$\begin{aligned} u_{k-k_u(k),t} &= (1 - f(\theta_{t-1})) u_{k,t-1} + \delta_t e_{k-k_u(k),t} \text{ for all } k \\ e_{k-k_u(k),t} &= (1 - \delta_{t-1}) e_{k-k_u(k)-1,t-1} + f(\theta_{t-1}) u_{k,t-1} \text{ for all } k. \end{aligned} \quad (18)$$

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

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

$$W_{k,t} = w_{k,t} + \beta \mathbb{E}_t \left[(1 - \delta_{t+1}) W_{k+1,t+1} + \delta_{t+1} U_{k+1,t+1} \right], \quad (19)$$

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

The value of an unemployed worker of type k is

$$U_{k,t} = b + \beta \mathbb{E}_t \left[f(\theta_t) (1 - \delta_{t+1}) W_{k-k_u(k),t+1} + (1 - f(\theta_t) (1 - \delta_{t+1})) U_{k-k_u(k),t+1} \right], \quad (20)$$

where b is the flow value of unemployment.¹⁶ If the worker contacts a firm and does not

¹⁶We will interpret b , interchangeably, as unemployment insurance since extending the model with a

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.

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

$$V_t = -\kappa + q(\theta_t)\beta\mathbb{E}_t \left[(1 - \delta_{t+1}) \sum_k \frac{u_{k,t}}{u_t} J_{k-k_u(k),t+1} + \delta_{t+1} V_{t+1} \right], \quad (21)$$

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

A firm that employs a worker of type k has value

$$J_{k,t} = p_{k,t} - w_{k,t} + \beta\mathbb{E}_t \left[(1 - \delta_{t+1}) J_{k+1,t+1} + \delta_{t+1} V_{t+1} \right], \quad (22)$$

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

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

$$S_{k,t} = W_{k,t} - U_{k,t} + J_{k,t} - V_t. \quad (23)$$

The individual value functions in Equations (19)–(22) and the definition of surplus in Equation (23) yield the following surplus value:

$$S_{k,t} = p_{k,t} - b + \beta\mathbb{E}_t \left[(1 - \delta_{t+1}) S_{k+1,t+1} - f(\theta_t)(1 - \delta_{t+1})\phi S_{k-k_u(k),t+1} + U_{k+1,t+1} - U_{k-k_u(k),t+1} \right]. \quad (24)$$

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

$$U_{k,t} = b + \beta\mathbb{E}_t \left[f(\theta_t)(1 - \delta_{t+1})\phi S_{k-k_u(k),t+1} + U_{k-k_u(k),t+1} \right]. \quad (25)$$

government levying lump-sum taxes to finance such a policy leaves the rest of the model unchanged.

The wage for worker type k is determined according to generalized Nash bargaining:

$$w_{k,t} = \arg \max (W_{k,t} - U_{k,t})^\phi (J_{k,t} - V_t)^{1-\phi}, \quad (26)$$

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

$$W_{k,t} - U_{k,t} = \phi S_{k,t} \quad \text{and} \quad J_{k,t} - V_t = (1 - \phi) S_{k,t}. \quad (27)$$

In words, the worker captures a constant share ϕ of the total match surplus, and the firm captures the rest.

Free entry of firms pins down $V_t = 0$ for all t . Equation (21) therefore implies

$$\frac{\kappa}{q(\theta_t)} = \beta \mathbb{E}_t \left[(1 - \delta_{t+1}) \sum_k \frac{u_{k,t}}{u_t} (1 - \phi) S_{k-k_u(k),t+1} \right]. \quad (28)$$

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

- match surplus S_k solves the Bellman equation in Equation (24) for all k ,
- unemployment value U_k solves the Bellman equation in Equation (25) for all k ,
- intermediate goods prices p_k satisfy Equation (17) for all k ,
- masses of (un)employed, u_k and e_k , follow the laws of motion in Equation (18) for all k ,
- end-of-period intermediate goods are given by $n_k = (1 - \delta)e_k$ for all k ,
- market tightness θ solves the free-entry condition in Equation (28),
- exogenous state variables z and δ 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 E 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 σ , and the relative weights of different types in production, α_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 1600 to extract the cyclical component of simulated time series.

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 $\mu = 0.72$, which falls well within the range of empirical estimates (see e.g., Petrongolo and Pissarides, 2001). Matching efficiency parameter \bar{m} is set such that the model matches the average US empirical quarterly job finding rate of 0.57.¹⁷ We impose the Hosios condition and set the bargaining power of workers equal to the elasticity of the matching function, $\phi = \mu$.¹⁸ Finally, the vacancy posting cost, κ , is set such that labor market tightness is normalized to $\theta = 1$ in steady state.

¹⁷To be consistent with our discrete time model, this empirical transition rate is not adjusted for time aggregation. Appendix C reports our main time series for UE flows from using time-aggregation-adjusted transition rates (i.e., permitting within-period EUE flows).

¹⁸The 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, a special case of the generalized Hosios condition derived in Mangin and Julien (2020).

The flow value of unemployment, $b = 0.39$, is set such that the replacement rate (relative to the average wage) is 40 percent, as in Shimer (2005). 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 σ) by setting the type-specific productivity weights α_k such that $p_k = 1$ for all k in steady state. We report details on this procedure in Appendix F.

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.

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 of firm-specific skills or the presence of a job ladder (see e.g. Jarosch, 2015; Jung and Kuhn, 2018).¹⁹

We show robustness to an extreme alternative downgrading specification below in Section 4.3, 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.

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

$$\ln(z_{t+1}) = (1 - \rho_z) \ln(\bar{z}) + \rho_z \ln(z_t) + \sigma_z \varepsilon_{t+1}^z \quad (29)$$

$$\ln(\delta_{t+1}) = (1 - \rho_\delta) \ln(\bar{\delta}) + \rho_\delta \ln(\delta_t) + \sigma_\delta \varepsilon_{t+1}^\delta, \quad (30)$$

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

¹⁹Alternatively, turbulence models permit gradual skill decline, but in these models all worker skill types are perfect substitutes in production (see e.g., Ljungqvist and Sargent, 1998, 2004; den Haan, Haefke, and Ramey, 2005).

Table 1: Parameter Values

Parameter	Value		Target/Source	Data	Model
	Congestion	No congestion			
β	0.99		Annual interest rate	0.04	0.04
μ	0.72		Shimer (2005)	0.72	0.72
\bar{m}	0.57		Job finding probability	0.57	0.57
η	0.72		Hosios condition	0.72	0.72
b	0.39		Avg. replacement rate	0.40	0.40
κ	0.21		Normalization $\theta = 1$	-	1.00
\bar{z}	1		Normalization	-	1.00
σ_z	0.008		St. dev. of ALP	0.010	0.010
ρ_z	0.956		Persistence of ALP	0.74	0.69
$\bar{\delta}$	0.037		Unemployment rate	0.063	0.063
σ_δ	0.107	0.080	St. dev. of UE/E	0.067	0.067
ρ_δ	0.709	0.670	Persistence of UE/E	0.84	0.74
$\rho_{\delta,z}$	-0.505	-0.560	corr(ALP, δ)	-0.41	-0.41
σ	0.241	1	Impulse response of θ to δ , see Figure 9		
α_k	see Appendix F	1	$p_k = 1$ for all k in steady state		

Note: Parameter values for both the baseline “congestion” model ($\sigma = 0.241$) and for the standard “no-congestion” model ($\sigma = 1$). 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.

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.²⁰ Finally, we let the correlation between ε_z and ε_δ be such that the model matches the correlation between average labor productivity and the separation rate observed in the data. We parameterize the aggregate shock processes jointly with the congestion parameter σ , which we describe below, as the behavior of UE flows is an equilibrium outcome.

4.2 Disciplining Congestion Parameter σ

Congestion is guided by the parameter that governs the elasticity of substitution between worker types, σ , which determines the degree of diminishing returns to specific worker types. We parameterize σ (jointly with 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, as the blue solid line, plots how this RMSE varies with the congestion parameter, σ . 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.

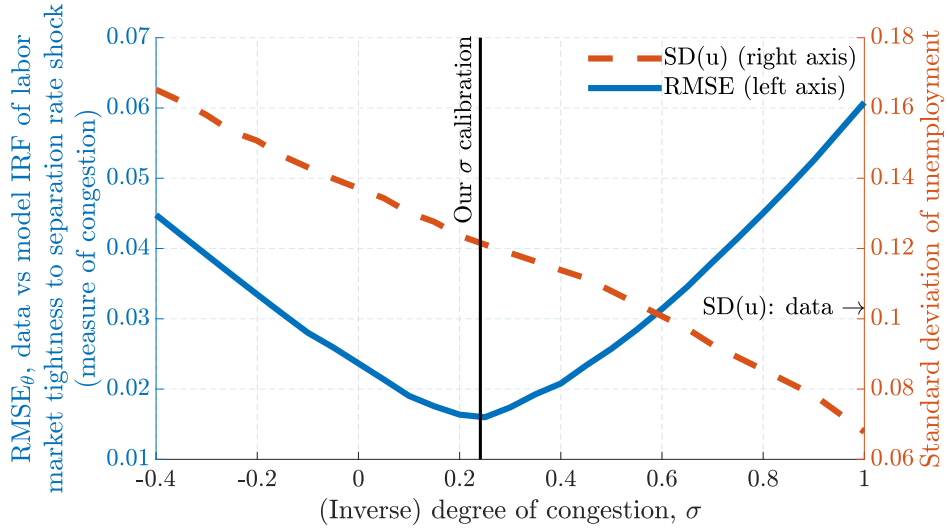
Figure 9 Panel (a) plots the IRF of labor market tightness to a separation shock in the calibrated model, with $\sigma = 0.241$, along with the empirical IRF. The model matches the empirical pattern well. Besides capturing the large negative impact response, the model also generates the observed persistent, 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 G, 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 (Appendix Figure A9).

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

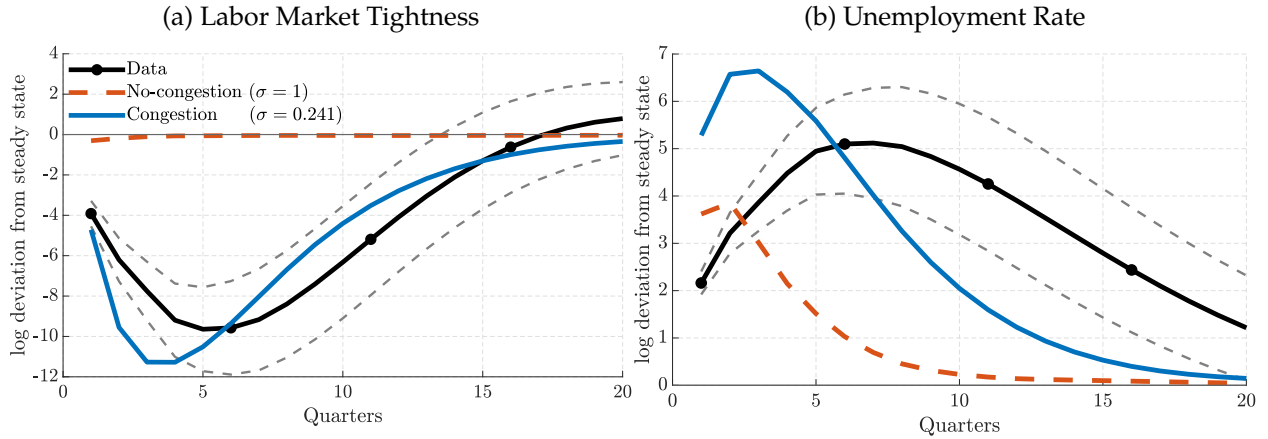
²⁰We target UE flows as a share in employment because of their key role for our congestion channel. We discuss alternative calibration approaches below.

Figure 8: Calibrating Congestion Parameter σ



Notes: 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.

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 H, 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 4.6 when studying overall unemployment fluctuations in the model.

4.3 Robustness to Alternative Congestion Mechanisms

Before moving to the quantitative analysis, we argue that alternative model structures that generate congestion will lead the model to behave very similarly if not identically 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 calibration is based on the parsimonious specification of the skill process described above: job loss resets worker types to $k = 1$, akin to models of the job ladder or firm-specific skill accumulation; the resulting concentration in one type generates a transparent notion of congestion in the CES production function. Yet in reality, a fair share of unemployed workers may reenter employment in their original type. For instance, some workers may not lose any skill, may be hired directly into higher-level positions, or be recalled.

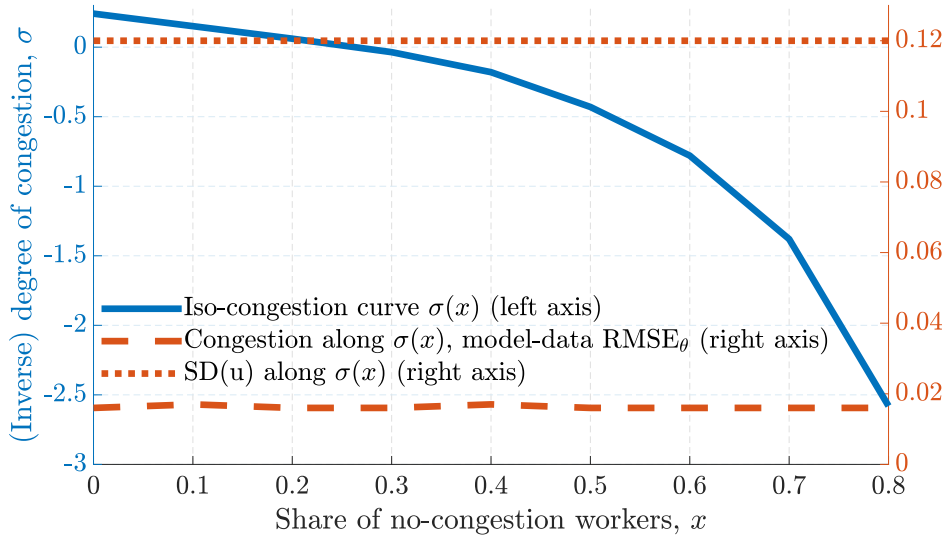
At first glance, such departures from our baseline specification would indeed seem to reduce the degree of congestion in hiring. However, for such model variants to still match the empirical degree of congestion, our calibration strategy simply would estimate a lower σ parameter for the relevant hiring margin subject to congestion. In the end, any such model variant will therefore exhibit the same degree of congestion.

To demonstrate robustness, we present an extreme alternative of the type evolution in Appendix I, and summarize it here. In this variant, fraction x of hires perfectly replicate the skill structure prevailing at the point of hiring, and can be thought of as “no-congestion” hires. Fraction $1 - x$ of hires continue to fully downgrade to $k = 1$. Isomorphically, the no-congestion workers operate in a separate linear production function, in which workers (formally, the intermediate inputs they produce) are perfect substitutes.²¹

Figure 10 shows the model properties for different values of x . $x = 0$ is exactly our original model with full downgrading to $k = 1$ upon job loss for all workers. Importantly,

²¹Here, the α_k -skill weights are recalibrated to yield homogeneous productivities in steady state. In this second interpretation, the final good is produced as a convex combination of the congestion (CRS-CES) and

Figure 10: Robustness to Alternative Specifications of Skill Process



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

each σ -model is reparameterized to match all the calibration targets. The only moment that a model with a fixed σ would miss is the impulse response of labor market tightness to separation rate shocks, which we used to pin down σ in the model with $x = 0$. As x increases, the figure illustrates that indeed, the repeatedly reparameterized model continues to hit the same RMSE target as for $x = 0$ (red dashed line). To achieve this fit, the extended model simply requires a lower and lower σ . Intuitively, as the share of new hires that do not create congestion increases, i.e., as x goes up, σ needs to fall in order to continue to match the negative IRF of labor market tightness to a separation shock. We plot the resulting “iso-congestion” $\sigma(x)$ curve as a function of x with the blue solid line. Appendix I derives this iso-congestion curve *analytically* as a function of x .²²

a no-congestion (linear) production functions:

$$Y = z \left[(1-x) \left(\sum_{k=1}^K \alpha_k^c (n_k^c)^\sigma \right)^{1/\sigma} + x \left(\sum_{k=1}^K \alpha_k^{nc} n_k^{nc} \right) \right], \quad (31)$$

where subscripts c and nc stand for the “congestion” and “no-congestion” sectors.

²²There, we consider a simple analytical expression for the elasticity of the marginal product of an average new hire \bar{p}_1 as a function of cohort size n_1 : $\varepsilon_{\bar{p}_1, n_1} = (\sigma - 1)(1 - n_1/N)(1 - x)$. The iso-congestion curve for a desired degree of congestion $\bar{\varepsilon}$ as a function of no-congestion worker share x is given by $\sigma(x, \bar{\varepsilon}) = 1 + \frac{\bar{\varepsilon}}{(1-x)(1-n_1/N)}$.

The figure also conveys the fact that the model properties are unchanged. The dotted red line shows that the standard deviation of unemployment remains the same for different values of x along the iso-congestion curve for σ as a function of x . Hence, with regards to the aggregate labor market quantities of alternative specifications of the skill process are *isomorphic* to our baseline specification in which all workers fall to $k = 1$ upon job loss, due to the recalibration of σ . Appendix I provides further details and formally shows this isomorphism to the baseline specification.

Congestion Through Convex Hiring Costs. In addition, in Appendix J 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 in a countercyclical way, in contrast to the procyclical DMP recruitment costs, which lead to dampening rather than amplification (as explained in, e.g., Shimer, 2010).²³ Since the model with convex hiring costs cannot speak to some of the additional macroeconomic applications we study in Section 5, our main specification obtains congestion from the production function.

4.4 A Bird’s Eye View of Business Cycle Statistics

We now study the quantitative implications of countercyclical congestion for labor market fluctuations, and dissect its amplification and propagation mechanisms. 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 cyclicity. Hence, countercyclical congestion can be viewed as a solution to the inability of the standard DMP model to generate realistic fluctuations in unemployment and labor market tightness (Shimer, 2005).

Robustness: Alternative Specifications. Appendix K replicates the analog of 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 G provides these statistics for the no-congestion model under the Hagedorn and Manovskii (2008) calibration, i.e., featuring a small match surplus in steady state

²³See, e.g., Fujita and Ramey (2007); Coles and Moghaddasi Kelishomi (2018) for models that relax the free entry condition along those lines.

Table 2: Business Cycle Properties: Data and Congestion Model

	ALP	f	δ	u	v	θ	UE/E	p_1
<i>Panel A: Data</i>								
Standard deviation	0.010	0.053	0.067	0.103	0.126	0.229	0.067	NA
Autocorrelation	0.746	0.871	0.773	0.934	0.926	0.936	0.836	NA
Correlation matrix								
ALP	1							
f	0.042	1						
δ	-0.415	-0.715	1					
u	-0.112	-0.931	0.848	1				
v	0.309	0.874	-0.869	-0.934	1			
θ	0.223	0.917	-0.874	-0.980	0.986	1		
UE/E	0.173	-0.722	0.567	0.833	-0.711	-0.783	1	
<i>Panel B: Congestion Model</i>								
Standard deviation	0.010	0.059	0.122	0.121	0.102	0.207	0.067	0.055
Autocorrelation	0.688	0.897	0.530	0.836	0.857	0.897	0.742	0.771
Correlation matrix								
ALP	1							
f	0.443	1						
δ	-0.410	-0.509	1					
u	-0.463	-0.924	0.743	1				
v	0.348	0.922	-0.157	-0.716	1			
θ	0.443	0.996	-0.514	-0.940	0.909	1		
UE/E	-0.337	-0.930	0.392	0.865	-0.876	-0.940	1	
p_1	0.490	0.952	-0.431	-0.862	0.900	0.949	-0.973	1

Notes: ALP , f , δ , u , θ , UE/E and p_1 indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, 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 1600.

(high b compared to productivity), which permits productivity shocks to have a larger effect on hiring and generate realistic labor market volatility. Appendix J 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 Michailat (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 δ compared to the data, but with lower autocorrelation. This is because we calibrate the sepa-

ration rate process to match the UE flows in the data rather than EU flows. We discuss this choice in Appendix H, 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.

4.5 Beveridge Curves

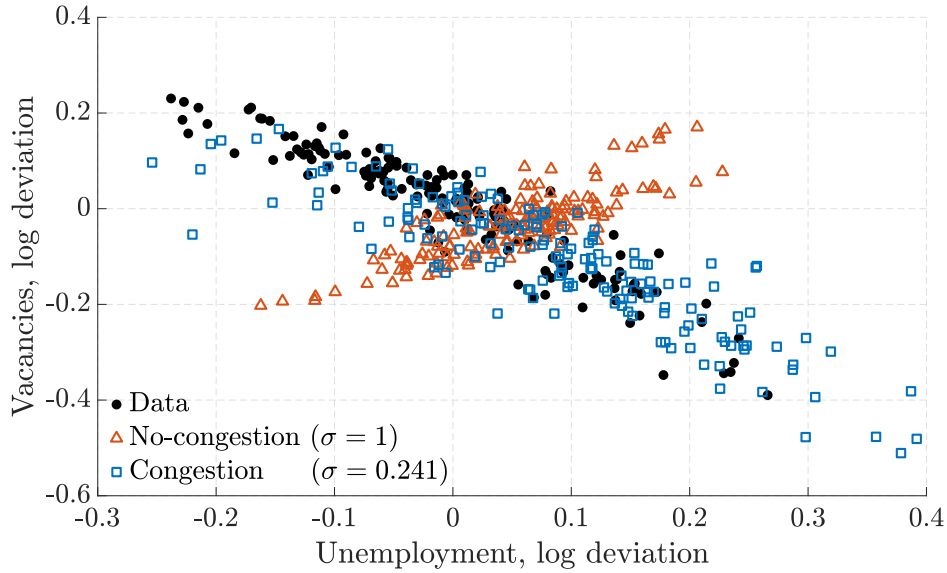
We now study the Beveridge curve, the relationship between vacancies and unemployment, and a crucial property of DMP models (see Elsby, Michaels, and Ratner, 2015, for a review). In fact, the Beveridge curve highlights the core difference between congestion and no-congestion models.

Figure 11 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. That is, in recessions, vacancies fall and the unemployment rate increases.

The no-congestion model features a counterfactually positive slope: on average, in recessions 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 in this model, which exhibits the Shimer (2005) puzzle. Instead, fluctuations in unemployment are largely due to separation shocks in this model. In response to separation shocks, the no-congestion model by its nature carries over its counterfactually flat IRF of labor market tightness to a separation rate shock, as described in Sections 2.2 and 4.2. That is, when separation shocks shift unemployment while not affecting TFP, firms post vacancies to keep labor market tightness stable, i.e., recessions are good times to hire such that firms post more vacancies when unemployment is high. 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

Figure 11: Beveridge Curve: Data and Models



Notes: The figure is a scatter plot of quarterly time series of vacancies (normalized by the labor force), based on the vacancy time series by (the Composite Help Wanted Index by Barnichon, 2010), against the unemployment rate in the U.S. data from 1976 to 2016, and the simulated time series from the no-congestion and congestion models. All variables are logged and HP-filtered using a smoothing parameter of 1600.

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.²⁴

4.6 The Volatility of Unemployment

Figure 8 visualizes the connection between the degree of congestion and the model's performance. Besides the calibration target discussed in Section 4.2 minimizing the RMSE between the model and empirical IRFs, it also plots, in red dashed lines, the volatility of unemployment for different values of σ (while recalibrating all other parameters to match the remaining targets). Countercyclical congestion, in the form of a higher degree of congestion (i.e., lower σ), amplifies unemployment volatility, the sources of which we discuss in the following section.

To build intuition for the amplification mechanisms in the model, consider a recession. As separations increase, unemployment rises. As long as the job finding rate does not decrease too much (due to a negatively correlated TFP shock and/or due to congestion),

²⁴Coles 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 (2015) for a discussion.

UE flows rise. The resulting increase in the share of new hires lowers their type-specific marginal product of labor, as long as $\sigma < 1$. The last column of Table 2 Panel B documents that the productivity of new hires is much more volatile and procyclical than average labor productivity. This drop in the productivity of new hires further reduces the incentives to hire, keeping unemployment elevated. Moreover, as new hires move together as a single cohort over the course of their employment spells, their abundance and hence depressed, cohort-specific productivity is persistent.

Importantly, any difference in amplification and propagation in our model relative to the standard framework is exclusively due to countercyclical congestion, i.e., the degree to which shifts in the employment share of new hires diminish their productivity. The crucial calibration choice that permits us to surgically isolate the role of countercyclical congestion in amplification is that we devise all model variants to feature the same, high fundamental match surplus by recalibrating the productivity weights α_k to generate the common unit productivity in steady state for all types, as described in Section 4.1.

4.7 Sources of Amplification: Productivity and Cohort Dynamics

The key to understanding amplification is the behavior of the match surplus for new hires, which enters the labor market clearing condition. Using Equation (24) and imposing the assumption that $k_u(k) = k - 1$ (i.e., full type downgrade), we can simplify the surplus expression for any worker type k as

$$S_{k,t} = \underbrace{p_{k,t}}_{\text{Current productivity}} - b + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) S_{k+1,t+1} \right]}_{\text{Continuation value at } k+1} - \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) f(\theta_t) \phi S_{1,t+1} \right]}_{\text{Worker's outside option: finding new job next period at } k=1}. \quad (32)$$

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 (32) to explicitly highlight these

three amplification channels, focusing on the surplus of new hires $k = 1$:

$$\begin{aligned}
S_{1,t} = & \underbrace{z_t - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t^s)) S_{t+1}^s \right]}_{\text{(i) No-congestion model surplus}} + \underbrace{S_t^* - S_t^s}_{\text{(ii) Flow productivity channel}} \\
& + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t)) \phi (S_{2,t+1} - S_{t+1}^*) \right]}_{\text{(iii) Present value channel}} + \underbrace{\beta \mathbb{E}_t \left[(1 - \delta_{t+1}) f(\theta_t) \phi (S_{2,t+1} - S_{1,t+1}) \right]}_{\text{(iv) Outside option channel}},
\end{aligned} \tag{33}$$

where $S_t^s = z_t - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t^s)) \phi S_{t+1}^s \right]$ is the surplus in the standard model without congestion and homogeneous workers, and θ^s is the associated labor market tightness.²⁵ $S_t^* = p_{1,t} - b + \beta \mathbb{E}_t \left[(1 - \delta_{t+1})(1 - f(\theta_t)) \phi S_{t+1}^* \right]$ is the match surplus in which flow productivity is (counterfactually) always equal to that of new hires, $p_{1,t}$. We now investigate these sources of amplification in detail.

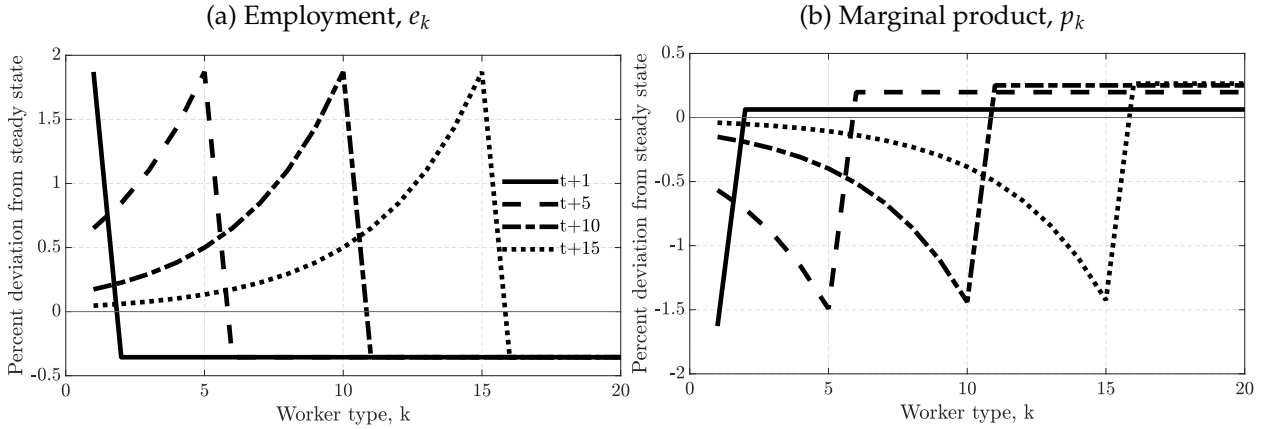
Flow Productivity Channel. As foreshadowed in Figure 7, Table 2 shows that counter-cyclical congestion dramatically amplifies the productivity of new hires, which is around five times as volatile as—and masked by the smoothness of—average productivity. It is also more procyclical, with a correlation with unemployment of -0.862 , compared to -0.463 for average productivity. Intuitively, in response to a positive separation shock, unemployment expands, UE flows rise, 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 of their employment spell but instead, persistent *cohort effects* arise as new hires are stuck with their initial cohort size as they move up the rungs of the type ladder together.

Figure 12 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 following the shock. For instance, the solid line, denoted by $t + 1$, shows the response for workers newly hired in the period following the aggregate separation shock. Because of the inflow of new hires, the mass of workers at the lowest type $k = 1$ expands (Panel (a)). Due to imperfect substitutability, this abundance pushes down their productivity (Panel

²⁵That is, for this standard surplus term, we use the counterfactually smooth job finding rate generated by the standard model to construct the standard surplus. All other terms use the same job finding rate generated by the congestion model.

Figure 12: Impulse Responses by Type k : Employment and Productivity



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

(b)). As time passes, the spikes in employment and productivity 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.²⁶ As a result of these persistent cohort effects, the expected present value of productivity of newly hired workers (formally, $\mathbb{E}_t \sum_{j=0}^{\infty} \beta^j (1 - \delta_{t+1+j}) p_{1+j,t+1+j}$)—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 our congestion model, a new hire starting an employment spell at step $k = 1$ in period t has productivity $p_{k=2,t+1}$ next period. By contrast, a new hire starting employment (i.e., in step $k = 1$) in the subsequent period $t + 1$ has an initial productivity of $p_{k=1,t+1}$. The differential productivities of these two types depend on their relative abundance in period $t + 1$, and similarly for all future periods.

At the Nash bargaining stage, 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

²⁶The 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.

in recessions, lowering worker's outside option, and expanding surplus.

In contrast, in our model 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 would be predicted on the basis of merely comparing productivity differences.²⁷

Quantifying the Sources of Amplification. We now quantify the contributions of the three channels to amplification arising from countercyclical congestion. We feed in counterfactual surpluses from subsets of the four channels in Equation (33) into the free-entry condition in Equation (28), and 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.

4.8 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' 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.

²⁷This outside option mechanism works through outside options in Nash bargaining and would not be present in for example models with wage rigidity Shimer (2004); Hall (2005b) or with wage setting protocols that insulate wages from outside options (Hall and Milgrom, 2008; Jäger, Schoefer, Young, and Zweimüller, 2020).

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

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

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

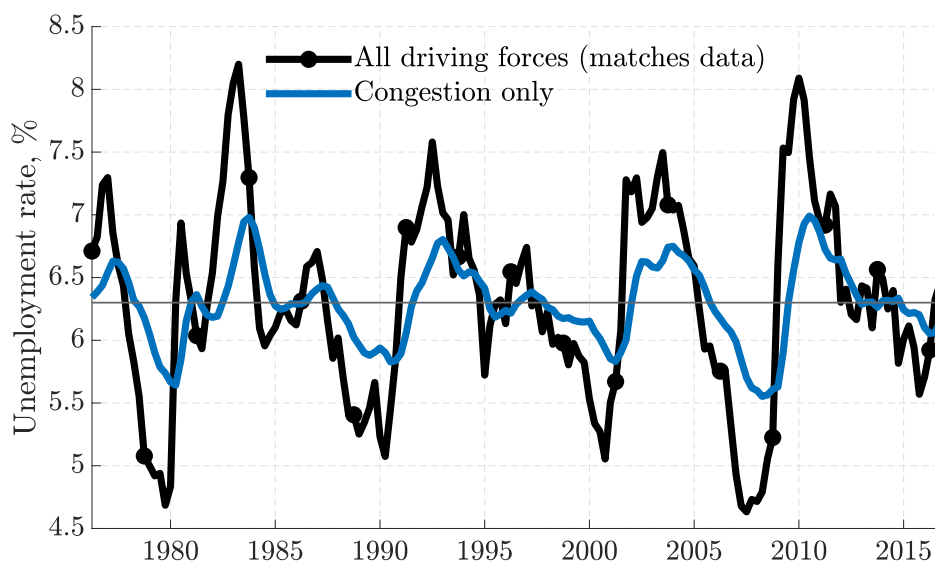
Method. Formally, we use the following equations for counterfactual unemployment, surplus, and labor market tightness that are purely driven by congestion:

$$\begin{aligned}
 u_{t+1}^c &= (1 - f(\theta_t^c))u_t^c + \bar{\delta}(1 - u_t^c) \\
 \kappa &= q(\theta_t^c)\beta\mathbb{E}_t(1 - \bar{\delta})S_{1,t}^c \\
 S_{k,t}^c &= p_{k,t} \cdot \frac{\bar{z}}{z_t} - b + \beta\mathbb{E}_t(1 - \bar{\delta})S_{k+1,t+1}^c - \beta\mathbb{E}_t(1 - \bar{\delta})f(\theta_t^c)\phi S_{1,t+1}^c \text{ for all } k.
 \end{aligned} \tag{34}$$

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 firm 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 U.S. time series data for average labor productivity and the share of new hires in employment (logged and HP-filtered with a smoothing parameter of 1600). Appendix Figure A15 presents both the estimated and empirical time series, Appendix E contains further details on the estimation procedure. Appendix L provides additional details on the decomposition, and additionally applies the method to TFP-only and separation-only counterfactuals.

Figure 13: Historical Decomposition of Unemployment: Actual and Congestion



Notes: The figure plots actual and congestion unemployment (u^c) estimated using data on the cyclical components of average labor productivity and new hires as a share of employment. The counterfactual unemployment time series for u^c is constructed based on the set of Equations (34).

The Time Series of Congestion-Driven Unemployment. Figure 13 shows the time path of unemployment implied by our model (which, as shown in Appendix Figure A15, essentially perfectly tracks the empirical time series) and congestion unemployment in the US. The figure shows that congestion alone is a powerful driver of unemployment fluctuations. The expansion of the employment share of new hires pushes down their productivity during recessions, and drives up unemployment. Conversely, during booms, new hires become scarce, which raises their productivity and pushes down unemployment. Quantitatively, the standard deviation of congestion-only unemployment is 0.05, about 40 percent that of overall unemployment. Computing the importance of congestion-only unemployment for unemployment variation as $cov(u, u^c)/var(u)$ reveals a 0.297 contribution. Therefore, countercyclical congestion explains between 30 and 40 percent of observed unemployment fluctuations.

Moreover, congestion-only unemployment moves closely with overall unemployment with a correlation coefficient between the two time series of 0.723. Finally, congestion-only unemployment is a source of persistence of overall unemployment—the autocorrelation coefficient of 0.950 relative to 0.905 of overall unemployment.

5 Additional Implications of Countercyclical Congestion

Besides providing a new perspective on unemployment fluctuations, countercyclical congestion rationalizes three additional macroeconomic patterns simultaneously: 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 such as unemployment insurance (UI) generosity. These applications can also be thought of as providing additional, external validation of our congestion model.

5.1 Business Cycle Accounting: The 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 equals the marginal product of labor (MPL). Both of these terms, in turn, also equal to the market-clearing real wage.

The Standard Labor Wedge. In the data, the MRS and the MPL exhibit a strongly cyclical gap called the *labor wedge* (Chari, Kehoe, and McGrattan, 2007; Shimer, 2009). This time-varying tax-like wedge, $1 - \tau_t$, is obtained as a residual—by specifying a utility function and an aggregate production function, and feeding in the empirical time series on consumption C , output Y , and employment E —from the following equation:

$$(1 - \tau) \cdot MPL = MRS \left(= \frac{-U_E(C, E)}{U_C(C, E)} \right). \quad (35)$$

Figure 14 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 η of extensive-margin labor supply $U(C, E) = \ln C - \Gamma E^{1+1/\eta}/(1+1/\eta)$. We set this elasticity to 0.34, as suggested by Chetty, Guren, Manoli, and Weber (2012).

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). Our framework offers such an explanation in the form of a more procyclical marginal product of labor, which is allocative for hiring.

Congestion and the Labor Wedge. To demonstrate how our model generates the labor wedge we, first, extend our aggregate production function to include capital, \tilde{K} , using a Cobb Douglas specification, with capital share a , and with the labor aggregator mirroring our baseline labor-based CES production function:

$$Y = z\tilde{K}^a \cdot \left(\left[\sum_{l=1}^K \alpha_l n_l^\sigma \right]^{\frac{1}{\sigma}} \right)^{1-a}. \quad (36)$$

Second, to retain comparability to the spot labor market structure (and equilibrium condition) in standard business cycle accounting, we sidestep the long-term nature of jobs and consider the productivity of new hires p_1 only. Recall, however, that we found in Section 4.7 that the present value of the productivity of new hires comoves tightly with flow productivity p_1 due to strong cohort effects.

To write out a spot labor market equilibrium condition, we reformulate the marginal product of new hires as the standard marginal labor product multiplicatively adjusted for by a term capturing diminishing returns to new hires, making clear that this term shows up *exactly* like the labor wedge in Equation (35):

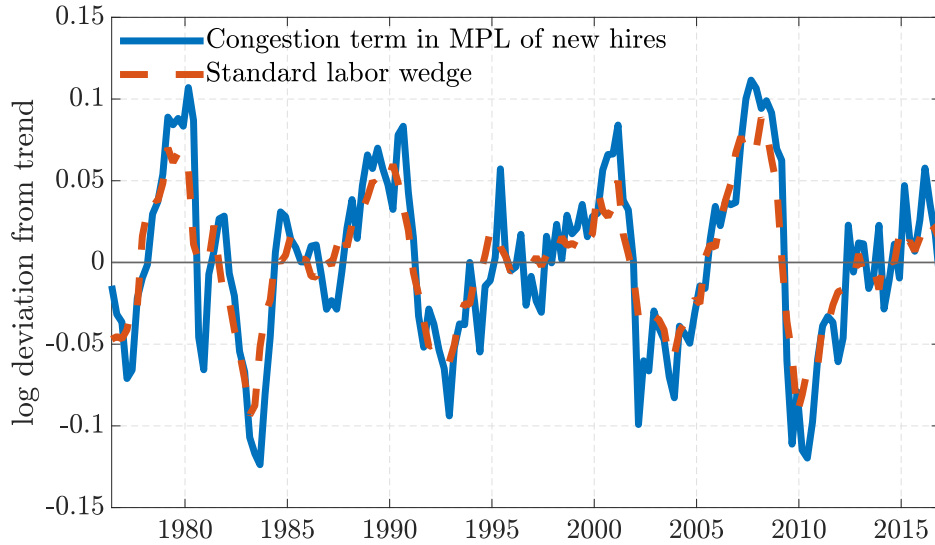
$$MRS = \underbrace{(1-a)\frac{Y}{N}}_{\text{Standard MPL}} \times \overbrace{\frac{\alpha_1 s_1^{\sigma-1}}{\sum_{l=1}^K \alpha_l s_l^\sigma}}^{MPL=p_1} \quad (37)$$

New-hire adjustment term

Figure 14 additionally plots the time series of this adjustment term for new hires' productivity (blue solid line), which strikingly closely tracks the standard labor wedge time series (correlation of 0.884).²⁸ That is, the economy with congestion provides an essentially full explanation of the labor wedge. Quantitatively, the remaining variation of the

²⁸We construct the new-hire term following the procedure for new-hire productivity in Figure 7 Panel (b), described in Footnote 14.

Figure 14: Standard Labor Wedge, and the Productivity of New Hires



Notes: The figure plots the labor wedge implied by the standard productivity measure and the wedge-like productivity adjustment term for new hires in Equation (37). All series are in logs and HP filtered using a smoothing parameter of 1600.

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 ($R^2 = 0.872$) for the standard labor wedge to 0.081 ($R^2 = 0.111$) for the difference.²⁹

5.2 Countercyclical Earnings Losses From Job Displacement

We now show that countercyclical congestion is consistent with cyclical properties of the wage trajectories of new hires, which echo the cohort effects of our model.

The Cyclicity 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),

²⁹We have also explored the role of the degree of congestion by studying the RMSE between the labor wedge and the new-hire adjustment term for different values of σ . Our calibrated value of $\sigma = 0.241$ is very close to the minimum of this u-shaped curve, which would be attained around $\sigma = 0.5$. Hence, this exercise also serves as an external validation for the empirical realism of our independently calibrated degree of congestion.

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.

Countercyclical congestion depresses wages of newly hired cohorts in recessions, leading to countercyclical displacement costs. Their relative abundance pushes down their marginal product and hence, their wage. These cohort effects are persistent while reemployed, until the cohort “dies out” through turnover.

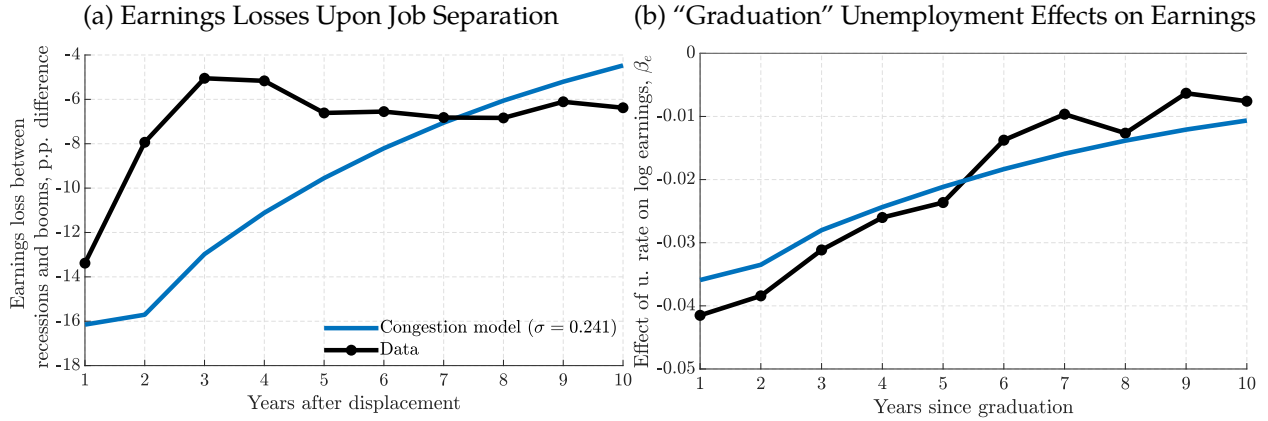
Figure 15 Panel (a) shows the *difference* in earnings effects from a job separation in recessions compared to booms for the model (blue solid line).³⁰ Workers displaced in a recession lose almost 15 percentage points more in earnings than workers displaced in booms. This difference fades only very gradually; 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.³¹ 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

³⁰Since our model is calibrated such that all worker types have identical wages in steady state ($p_k = 1$ and hence $w_k = w$ for all k), it cannot speak to the *level* of displacement costs.

³¹The 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 15 Panel (a).

Figure 15: Path Dependence of Earnings Losses



Notes: Panel (a) plots the percentage point difference in earnings losses of displaced workers in recessions relative to booms in the data (Davis and von Wachter, 2011, Figure 4 Panel C), and in the congestion model. Panel (b) plots the effect of the business cycle (unemployment rate) at point of graduation on earnings over time in the data (Schwandt and von Wachter, 2019, Figure 2) and in the model. The model results are based on estimating the regression specification in Equation (38) using simulations from our baseline model.

contain a life-cycle dimension, we can proxy for it in our model by following newly hired workers entering the labor market with type $k = 1$. We estimate the following regression on model-simulated earnings paths of cohorts of newly hired workers, which mimics Equation (2) estimated on data in Schwandt and von Wachter (2019):

$$y_{g,t} = \alpha + \beta_e u_g + \lambda_g + \chi_t + \epsilon_{g,t}, \quad (38)$$

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

Figure 15 Panel (b) plots the β_e coefficients estimated on simulated data together with the empirical estimates from Schwandt and von Wachter (2019).³² 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.

³²See Figure 2 in Schwandt and von Wachter (2019) for the empirical estimates in our Figure 15 Panel (b).

Mechanisms. Most of the proximate sources of these two types of countercyclical earnings losses are accounted for by wage declines conditional on working rather than increased nonemployment (see, e.g., von Wachter, forthcoming), supporting the persistent cohort effects on productivity in our model. The cohort effects also are associated with or can be accounted for by 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 low-quality relative to high-quality firms being more able to absorb the increase in UE hires. A complementary literature studies the destruction and creation of jobs by firm quality (Moscarini and Postel-Vinay, 2012; Haltiwanger, Hyatt, Kahn, and McEntarfer, 2018), and the countercyclicality of skill requirements (Modestino, Shoag, and Ballance, 2016).

Related Evidence: New-Hire Wage Cyclicalities. The higher degree of procyclicality of new hires' productivity is also consistent with new hires' wages being more procyclical than incumbents' wages (Pissarides, 2009). Yet, we refrain from quantitatively matching these patterns because first-period flow wages need not be informative about the present value of wages (Shimer, 2004), through which our cohort effects would show up and which the two exercises above capture.³³ These effects have been linked to job quality fluctuations (Gertler, Huckfeldt, and Trigari, 2020).

5.3 Policy Insensitivity Despite Productivity-Driven Business Cycles

We close our paper 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 policy that affect job surplus, such as unemployment insurance benefits. Costain and Reiter (2008) estimate the semi-elasticity of the unemployment rate with respect to the replacement rate, $\epsilon_{u,b/\bar{w}} = \partial \ln u / \partial (b/\bar{w})$, to lie between 2 and 3.5 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 that calibration dramatically overstates the semi-elasticity of unemployment with respect

³³As a notable exception, Kudlyak (2014) computes present values of wages in an intermediate step, but then reports only the cyclical behavior of the user cost of labor (a differenced version of wages across cohorts). She finds a large cyclicity of new hires' wages, consistent with our model with cohort effects.

to the replacement rate in the cross-country and long-run context (for empirical research on short run effects across US local labor markets, see Hagedorn, Karahan, Manovskii, and Mitman, 2019; Chodorow-Reich, Cogleianese, 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., $b_{new} = 1.01b_{base}$, 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 $\epsilon_{u,b/\bar{w}} = \frac{\ln u_{new} - \ln u_{base}}{(b_{new}/\bar{w}_{new}) - (b_{base}/\bar{w}_{base})} \approx 2.6$, a value well within the bounds reported by Costain and Reiter (2008). Hence, our framework *simultaneously* matches the high volatility of labor market variables and the lower sensitivity of these variables with respect to policy instruments. This is because our model generates labor market volatility through large fluctuations in allocative productivity (and the presence of cohort effects), and hence it can “afford” to keep the fundamental surplus (and hence the elasticity of unemployment to productivity) at the high value in Shimer (2005).

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 additionally being consistent with the relative insensitivity of labor market variables to labor market policies such as unemployment insurance, and, relatedly, while featuring a high fundamental surplus and not relying on wage rigidity. Our framework also rationalizes the countercyclical labor wedge, by ascribing it to a highly volatile allocative productivity, and the larger earnings losses upon job displacement or the lifetime earnings effects on cohorts of graduating during recessions.

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 supporting this congestion mechanism. A dedicated future test of

congestion would identify a pure disturbance in the number of the unemployed and study hiring responses at the aggregate level. Second, our study suggests a link between separation shocks and job finding rates, which existing statistical studies agnostically treat as independent regarding their contribution to unemployment fluctuations. 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 applications has supported our productivity-based congestion mechanism, we have shown that congestion may emerge also from hiring costs or perhaps other factors. Using detailed micro data to distinguish between specific sources of congestion that fall into the class of congestion mechanisms our paper proposes may be fruitful.

References

- ABRAHAM, K. AND L. KATZ (1986): "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?" *Journal of Political Economy*, 94, 507–522.
- ALTONJI, J. G., L. B. KAHN, AND J. D. SPEER (2016): "Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success," *Journal of Labor Economics*, 34, S361–S401.
- BARNICHON, R. (2010): "Building a Composite Help Wanted Index," *Economics Letters*, 109, 175–178.
- BILS, M., P. J. KLENOW, AND B. A. MALIN (2018): "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, 108, 1118–46.
- BLANCHARD, O. AND P. DIAMOND (1990): "The Cyclical Behavior of the Gross Flows of US Workers," *Brookings Papers on Economic Activity*, 1990, 85–155.
- BOONE, C., A. DUBE, L. GOODMAN, AND E. KAPLAN (forthcoming): "Unemployment insurance generosity and aggregate employment," *American Economic Journal: Policy*.
- BURDA, M. AND C. WYPLOSZ (1994): "Gross Worker and Job Flows in Europe," *European Economic Review*, 38, 1287–1315.
- CAHUC, P., S. CARCILLO, AND T. LE BARBANCHON (2018): "The Effectiveness of Hiring Credits," *The Review of Economic Studies*, 86, 593–626.
- CARD, D. AND T. LEMIEUX (2000): "Can Falling Supply Explain the Rising Return to College for Younger Men?" *Quarterly Journal of Economics*, 116, 705–746.
- CHARI, V., P. KEHOE, AND E. MCGRATTAN (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.
- CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2012): "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macro Annual 2012*.
- CHODOROW-REICH, G., J. COGLIANESE, AND L. KARABARBOUNIS (2019): "The Macro Effects of Unemployment Benefit Extensions: a Measurement Error Approach," *The Quarterly Journal of Economics*, 134, 227–279.
- CHODOROW-REICH, G. AND J. WIELAND (2020): "Secular Labor Reallocation and Business Cycles," *Journal of Political Economy*, 128, 2245–2287.
- CHRISTIANO, L., M. EICHENBAUM, AND M. TRABANDT (2016): "Unemployment and Business Cycles," *Econometrica*, 84, 1523–1569.
- (2020): "Why is Unemployment so Countercyclical?" *Working Paper*.
- COLES, M. AND A. MOGHADDASI KELISHOMI (2018): "Do Job Destruction Shocks Matter in the Theory of Unemployment?" *American Economic Journal: Macroeconomics*, 10, 118–36.
- COSTAIN, J. AND M. REITER (2008): "Business Cycles, Unemployment Insurance, and the Calibration of Matching Models," *Journal of Economic Dynamics and Control*, 32.
- DAVIS, S., J. FABERMAN, AND J. HALTIWANGER (2006): "The Flow Approach to Labor Markets: New Data Sources and Micro-Macro Links," *Journal of Economic Perspectives*, 20, 3–26.

- DAVIS, S. AND T. VON WACHTER (2011): "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*, 1–72, autumn.
- DEN HAAN, W., C. HAEFKE, AND G. RAMEY (2005): "Turbulence and Unemployment in a Job Matching Model," *Journal of the European Economic Association*, 3, 1360–1385.
- DEN HAAN, W., G. RAMEY, AND J. WATSON (2000): "Job Destruction and Propagation of Shocks," *American Economic Review*, 90, 482–498.
- DOERINGER, P. B. AND M. J. PIORE (1985): *Internal Labor Markets and Manpower Analysis*, ME Sharpe.
- DORAN, K., A. GELBER, AND A. ISEN (2020): "The Effects of High-Skilled Immigration Policy on Firms: Evidence from H-1B Visa Lotteries," *Working Paper*.
- DUPRAZ, S., E. NAKAMURA, AND J. STEINSSON (2019): "A Plucking Model of Business Cycles," *Working Paper*.
- ELSBY, M., R. MICHAELS, AND D. RATNER (2015): "The Beveridge Curve: A Survey," *Journal of Economic Literature*, 53, 571–630.
- ELSBY, M. W. L., B. HOBIJN, AND A. ŞAHİN (2013): "Unemployment Dynamics in the OECD," *The Review of Economics and Statistics*, 95, 530–548.
- ENGBOM, N. (2020): "Recruiting Talent During a Recession," *Working Paper*.
- FERRARO, D. (2018): "The Asymmetric Cyclical Behavior of the US Labor Market," *Review of Economic Dynamics*, 30, 145–162.
- FUJITA, S. AND G. MOSCARINI (2017): "Recall and Unemployment," *American Economic Review*, 107, 3875–3916.
- FUJITA, S. AND G. RAMEY (2006): "The Cyclicalities of Job Loss and Hiring," *Working Paper*.
- (2007): "Job Matching and Propagation," *Journal of Economic Dynamics and Control*, 31, 3671–3698.
- (2009): "The Cyclicalities of Separation and Job Finding Rates," *International Economic Review*, 50, 415–430.
- GATHMANN, C., I. HELM, AND U. SCHÖNBERG (2018): "Spillover Effects of Mass Layoffs," *Journal of the European Economic Association*, 18, 427–468.
- GERTLER, M., C. HUCKFELDT, AND A. TRIGARI (2020): "Unemployment Fluctuations, Match Quality and the Wage Cyclicalities of New Hires," *The Review of Economic Studies*.
- GERTLER, M. AND A. TRIGARI (2009): "Unemployment Fluctuations with Staggered Nash Wage Bargaining," *Journal of Political Economy*, 117, 38–86.
- GÓMEZ, V., A. MARAVALL, AND D. PEÑA (1999): "Missing Observations in ARIMA Models: Skipping Approach versus Additive Outlier Approach," *Journal of Econometrics*, 88, 341–363.
- HAGEDORN, M., F. KARAHAN, I. MANOVSKII, AND K. MITMAN (2019): "Unemployment Benefits and Unemployment in the Great Recession: the Role of Macro Effects," *NBER Working Paper*.
- HAGEDORN, M. AND I. MANOVSKII (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49.
- (2013): "Job Selection and Wages over the Business Cycle," *American Economic Review*, 103,

771–803.

- HALL, R. (2005a): “The Amplification of Unemployment Fluctuations through Self-Selection,” *NBER Working Paper*.
- (2005b): “Employment Fluctuations with Equilibrium Wage Stickiness,” *American Economic Review*, 95, 50–65.
- (2017): “High Discounts and High Unemployment,” *American Economic Review*, 107, 305–30.
- HALL, R. AND M. KUDLYAK (2020a): “The Inexorable Recoveries of US Unemployment,” *NBER Working Paper*.
- (2020b): “Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?” *NBER Working Paper*.
- HALL, R. AND P. MILGROM (2008): “The Limited Influence of Unemployment on the Wage Bargain,” *American Economic Review*, 98, 1653–74.
- HALTIWANGER, J. C., H. R. HYATT, L. B. KAHN, AND E. MCENTARFER (2018): “Cyclical Job Ladders by Firm Size and Firm Wage,” *American Economic Journal: Macroeconomics*, 10, 52–85.
- HUCKFELDT, C. (2016): “Understanding the Scarring Effect of Recessions,” *Working Paper*.
- JÄGER, S. AND J. HEINING (2019): “How Substitutable are Workers? Evidence from Worker Deaths,” *Working Paper*.
- JÄGER, S., B. SCHOEFFER, S. YOUNG, AND J. ZWEIMÜLLER (2020): “Wages and the Value of Nonemployment,” *The Quarterly Journal of Economics*, 135, 1905–1963.
- JAROSCH, G. (2015): “Search for Job Security and the Consequences of Job Loss,” *Working Paper*.
- JEONG, H., Y. KIM, AND I. MANOVSKII (2015): “Price of Experience,” *American Economic Review*, 105, 784–815.
- JUNG, P. AND M. KUHN (2018): “Earnings Losses and Job Mobility over the Lifecycle,” *Journal of the European Economic Association*, 17, 678 – 724.
- KAHN, L. (2014): “The Longterm Labor Market Consequences of Graduating College in a Bad Economy,” *Labour Economics*, 17, 206–223.
- KARABARBOUNIS, L. (2014): “The Labor Wedge: MRS vs MRN,” *Review of Economic Dynamics*, 17, 303–316.
- KATZ, L. AND B. MEYER (1990): “Unemployment Insurance, Recall Expectations, and Unemployment Outcomes,” *The Quarterly Journal of Economics*, 105, 973–1002.
- KATZ, L. AND K. MURPHY (1992): “Changes in Relative Wages, 1963-87: Supply and Demand Factors,” *Quarterly Journal of Economics*, 107, 35–78.
- KRUSELL, P., T. MUKOYAMA, R. ROGERSON, AND A. ŞAHİN (2017): “Gross Worker Flows Over the Business Cycle,” *American Economic Review*, 107, 3447–76.
- KRUSELL, P., L. OHANIAN, J.-V. RIOS-RULL, AND G. VIOLANTE (2000): “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, 68, 1029–1053.
- KUDLYAK, M. (2014): “The Cyclicalities of the User Cost of Labor,” *Journal of Monetary Economics*, 68, 53–67.

- LILIEN, D. (1982): "Sectoral Shifts and Cyclical Unemployment," *Journal of Political Economy*, 90, 777–793.
- LJUNGQVIST, L. AND T. SARGENT (1998): "The European Unemployment Dilemma," *Journal of Political Economy*, 106, 514–550.
- (2004): "European Unemployment and Turbulence Revisited in a Matching Model," *Journal of the European Economic Association*, 2, 456–468.
- (2017): "The Fundamental Surplus," *American Economic Review*, 107.
- MANGIN, S. AND B. JULIEN (2020): "Efficiency in Search and Matching Models: A Generalized Hosios Condition," *Working Paper*.
- MERCAN, Y. AND B. SCHOEFER (2020): "Jobs and Matches: Quits, Replacement Hiring, and Vacancy Chains," *American Economic Review: Insights*, 2, 101–24.
- MIAN, A. AND A. SUFI (2014): "What Explains the 2007–2009 Drop in Employment?" *Econometrica*, 82, 2197–2223.
- MICHAILLAT, P. (2012): "Do Matching Frictions Explain Unemployment? Not in Bad Times," *American Economic Review*, 102, 1721–50.
- MODESTINO, A. S., D. SHOAG, AND J. BALLANCE (2016): "Downskilling: Changes in Employer Skill Requirements over the Business Cycle," *Labour Economics*, 41, 333–347.
- MOSCARINI, G. AND F. POSTEL-VINAY (2012): "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment," *American Economic Review*, 102, 2509–39.
- MUELLER, A. (2017): "Separations, Sorting, and Cyclical Unemployment," *American Economic Review*, 107, 2081–2107.
- OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4, 1–29.
- PETRONGOLO, B. AND C. PISSARIDES (2001): "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39.
- PISSARIDES, C. A. (2009): "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica*, 77, 1339–1369.
- SCHMIEDER, J., T. VON WACHTER, AND J. HEINING (2019): "The Costs of Job Displacement over the Business Cycle and its Sources: Evidence from Germany," *Working Paper*.
- SCHOEFER, B. (2015): "The Financial Channel of Wage Rigidity," *Working Paper*.
- SCHWANDT, H. AND T. VON WACHTER (2019): "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets," *Journal of Labor Economics*, 37, S161–S198.
- SHIMER, R. (2004): "The Consequences of Rigid Wages in Search Models," *Journal of the European Economic Association*, 2, 469–479.
- (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49.
- (2009): "Convergence in Macroeconomics: The Labor Wedge," *American Economic Journal*:

- Macroeconomics*, 1, 280–97.
- (2010): *Labor Markets and Business Cycles*, Princeton University Press.
- (2012): “Reassessing the Ins and Outs of Unemployment,” *Review of Economic Dynamics*, 15, 127 – 148.
- SILVA, J. AND M. TOLEDO (2013): “The Unemployment Volatility Puzzle: The Role of Matching Costs Revisited,” *Economic Inquiry*, 51, 836–843.
- VON WACHTER, T. (forthcoming): “The Persistent Effects of Initial Labor Market Conditions for Young Adults and their Sources,” *Journal of Economic Perspectives*.

Online Appendix of: Countercyclical Congestion

Yusuf Mercan, Benjamin Schoefer and Petr Sedláček

Contents

A	Measuring Worker Flows	59
B	Evidence from OECD Countries	65
C	Discrete versus Time-Aggregation-Adjusted Data	67
D	A Generalization of the Baseline Model: Types vs. Inputs	70
E	Solution Method	71
F	Details of the Baseline Parameterization: Homogeneous Steady State Marginal Products Across Types	74
G	Alternative Calibration: Small Surplus/“High b”	75
H	Alternative Calibration: Matching EU Flows	78
I	Deriving the Iso-congestion Curve	82
J	Alternative Mechanism: Convex Hiring Costs	85
K	Business Cycle Statistics in the Standard Model	87
L	Historical Decomposition: Additional Material	89

List of Tables

A1	Discrete vs. Time-Aggregation Adjusted Worker Transitions	68
A2	Business Cycle Properties: No-Congestion, Low-Surplus Model	77
A3	Business Cycle Properties of the Congestion Model: Baseline and Alternative Calibration	81
A4	Business Cycle Properties: Convex Hiring Cost Model	86
A5	Business Cycle Properties of the No-Congestion Model	88
A6	Historical Decomposition of Unemployment: Model and Counterfactuals .	91

List of Figures

A1	Countercyclicality of the Employment Share with Nonemployment Past Year	61
A2	The Countercyclicality of New Hire Share: CPS Worker Flows	62
A3	Cyclicality of Share of New Hires in Employment: CPS Worker Flows	63
A4	Cyclicality of New Hires: CPS Worker Flows	64
A5	Cyclicality of UE Flows in the OECD	66
A6	Comparing Discrete and Time Aggregation Adjusted UE Shares	69
A7	Labor Productivity: Empirical Impulse Responses to a Separation Shock . .	72
A8	Relative Worker Productivities in the Congestion Model	74
A9	Impulse Responses to a Separation Shock: No-Congestion, Low-Surplus Model	76
A10	Separation Rate: Measured and Implied	79
A11	Impulse Responses to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows	80
A12	Impulse Responses of Separation Rate to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows . . .	80
A13	Iso-congestion Curves	84
A14	Impulse Responses to a Separation Shock: Convex Hiring Cost Model	86
A15	Time Paths of Labor Market Variables	90
A16	Unemployment Components	91

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 U.S. 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.

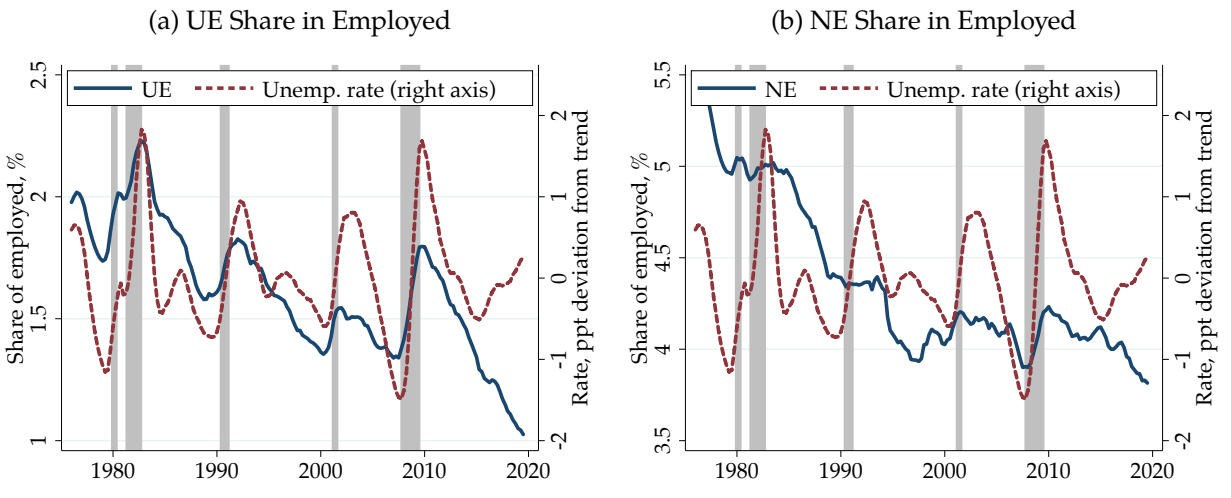
As a robustness check discussed in Section 2.1, below in Figures A1 to A4, we also replicate main text Figure 1 by considering the nonemployment (comprising unemployment and out of the labor force) rather than the unemployment history of the employed, and find qualitatively similar cyclical patterns. 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 4.2.

Figure A1: Countercyclicity of the Employment Share with Nonemployment Past Year



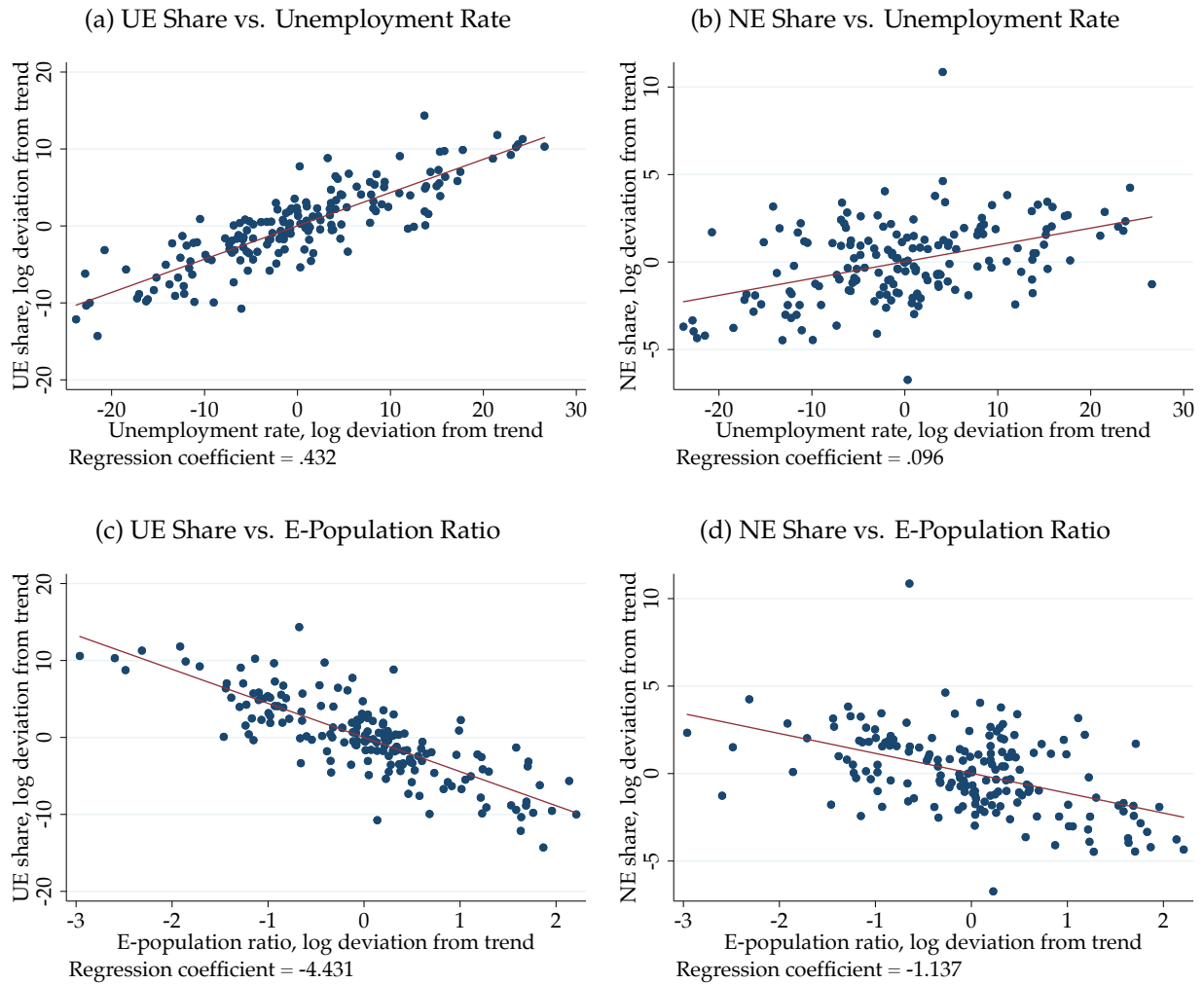
Notes: The figure replicates Figure 1, but instead conditions on *nonemployment* duration, i.e., we also include labor market states where a worker might be out of the labor force. Panel (a) plots the share of employed workers who have undergone a nonemployment spell in the preceding calendar year for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

Figure A2: The Countercyclicity of New Hire Share: CPS Worker Flows



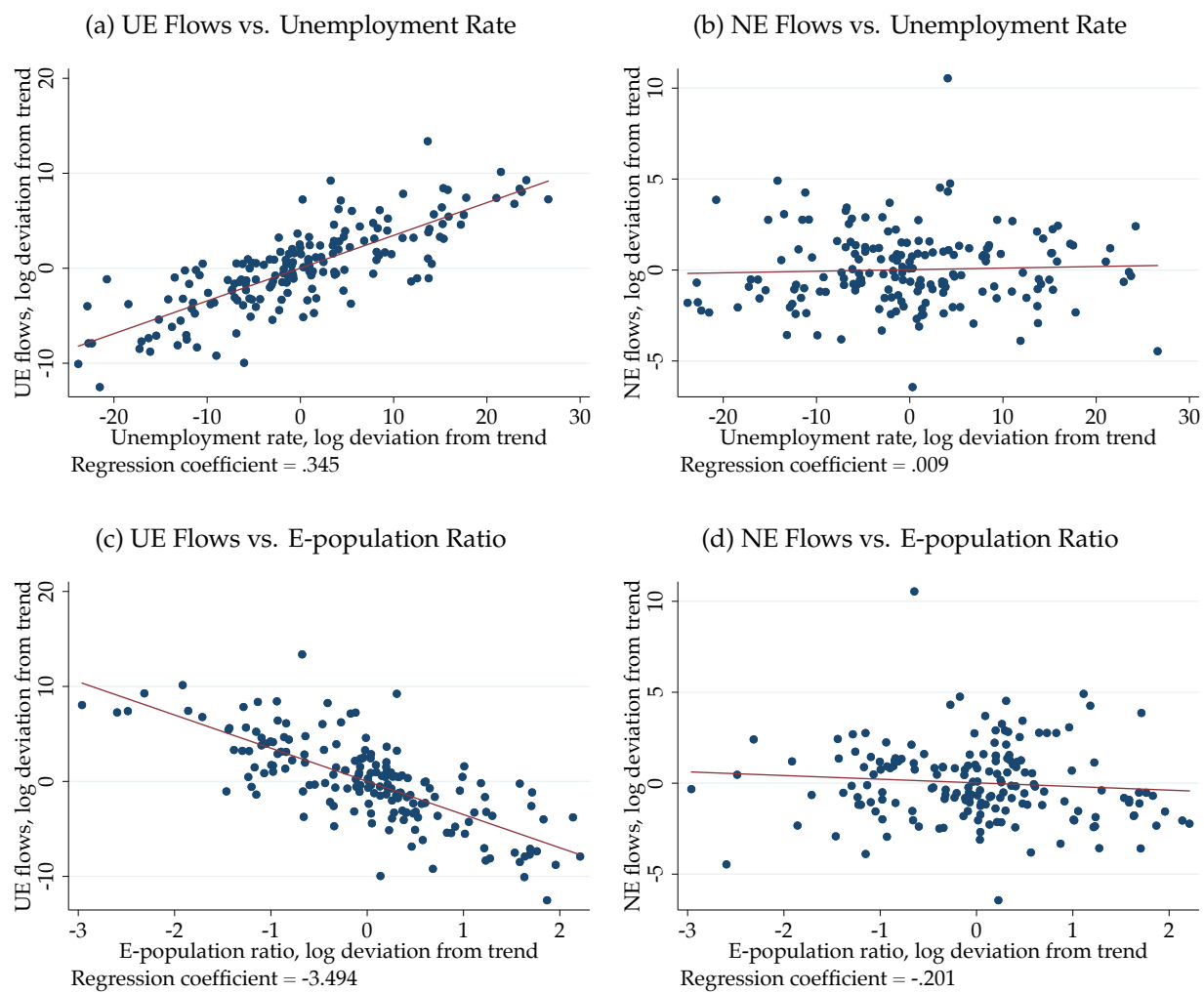
Notes: Panel (a) plots the share of UE hires in employment. Panel (b) plots NE flows in the share of employed. All time series are based on quarterly averages of monthly data and for visual clarity are smoothed by taking centered four-quarter moving averages. Both panels also plot the percentage point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

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



Notes: The figure plots different measures of new-hire share in employment (UE or NE) against employment measures (unemployment rate or employment-population ratio). All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1600. Source: CPS monthly files.

Figure A4: Cyclicality of New Hires: CPS Worker Flows



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 1600. While our model relies on 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.

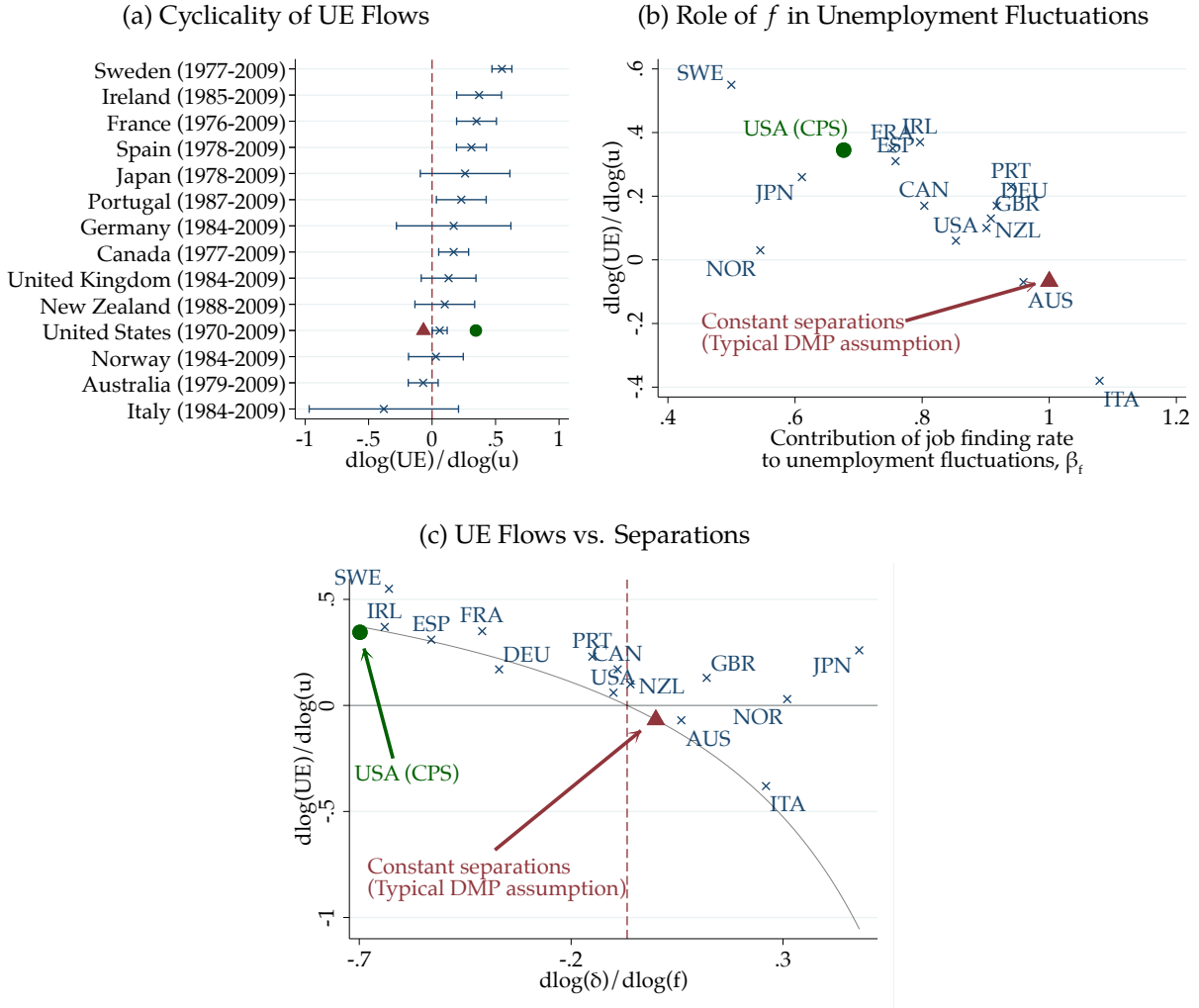
B Evidence from OECD Countries

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 $du^f/u^f = -(1-u)df/f$. Fujita and Ramey (2009) show that the regression coefficient of du^f/u^f on du/u also represents the share of the variance in unemployment rate fluctuations due to fluctuations in the job finding rate (rather than in the job separation rate). The smaller this share, the more countercyclical the UE flows on average, since $\frac{dUE/UE}{du/u} = -\frac{1}{1-u} \frac{du^f/u^f}{du/u} + 1$. Drawing on cross-country differences in the OECD, we document the empirical validity of this theoretical property in Panel (b) of Figure 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: Cyclicity of UE Flows in the OECD



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

C Discrete versus Time-Aggregation-Adjusted Data

Our baseline measure 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.

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 \widehat{f} and $\widehat{\delta}$ (which correspond to f and δ in the main text).

We start with the monthly job finding \widehat{f}_t and separation $\widehat{\delta}_t$ probabilities, as explained in Appendix A, underlying the analysis in the main text.

Second, we compute the monthly job finding and separation hazards, f_t and δ_t , solving the following system of equations:

$$\begin{aligned}\widehat{\delta}_t &= u_{ss,t}(1 - e^{-f_t - \delta_t}) \\ \widehat{f}_t &= (1 - u_{ss,t})(1 - e^{-f_t - \delta_t}),\end{aligned}\tag{A1}$$

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

$$U_{t-1+\tau} = \frac{(1 - e^{-(f_t + \delta_t)\tau})\delta_t}{f_t + \delta_t} L_{t-1} + e^{-(f_t + \delta_t)\tau} U_{t-1},\tag{A2}$$

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

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

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

Notes: The table compares the time series properties of UE flows based on our discrete time measurement approach used in the main text to a version corrected for time-aggregation bias. All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1600.

$t - 1 + \tau$ until t for $\tau \in [0, 1)$, conditional on having a job at t .¹ This probability is given by

$$\lim_{\Delta \rightarrow 0} (1 - \Delta \delta_t)^{\frac{1-\tau}{\Delta}} = e^{-\delta_t(1-\tau)}. \quad (\text{A3})$$

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

$$UE_t = \int_0^1 \underbrace{f_t}_{\text{Find job}} \underbrace{U_{t-1+\tau}}_{\text{Number of unemployed}} \underbrace{e^{-\delta_t(1-\tau)}}_{\text{Do not lose job}} d\tau. \quad (\text{A4})$$

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

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

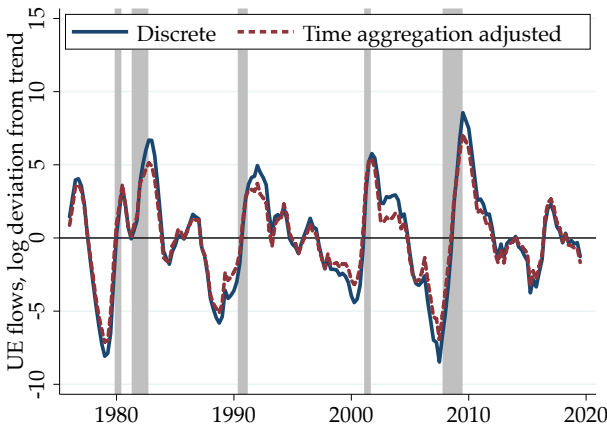
Table A1 summarizes the properties of the time series we use in the main text and the time series we construct using the alternative approach presented above. The two time series have extremely similar standard deviations and autocorrelations, and are nearly perfectly correlated.

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

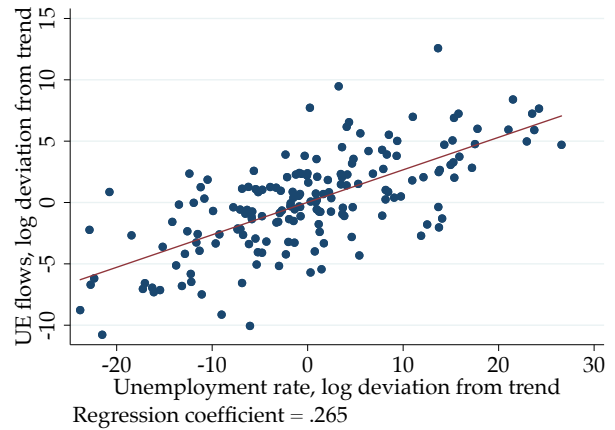
¹Therefore, our results do not study cycles such as "EUEUEUE" transitions during the period. These are comparatively tiny compared to the first-order flows stemming from the initially employed losing their job during the period, becoming reemployed, and not losing that first-found job again.

Figure A6: Comparing Discrete and Time Aggregation Adjusted UE Shares

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



(b) UE Flows vs. Unemployment Rate



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. All time series are based on quarterly averages of monthly data and are logged and HP-filtered using a smoothing parameter of 1600. Source: CPS monthly files.

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.

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

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

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

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

$$p_i = \alpha_i n_i^{\sigma-1} \frac{Y}{\sum_j \alpha_j n_j^\sigma} = \alpha_i s_i^{\sigma-1} \frac{1}{\sum_j \alpha_j s_j^\sigma} \frac{Y}{N}, \quad (\text{A6})$$

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 .

E 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.

E.1 Steady State

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

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

In addition, under our 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 (28). 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 α_k via

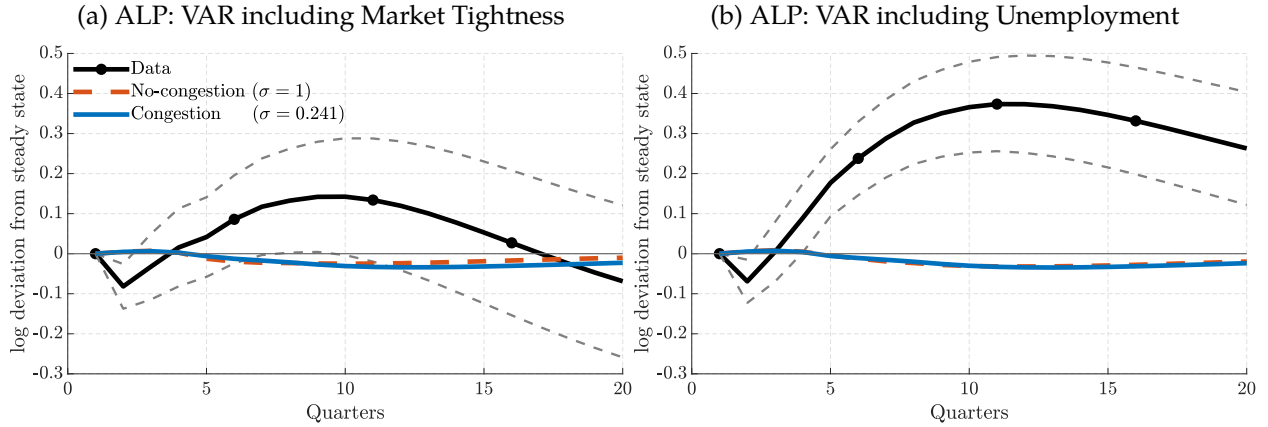
$$1 = p_k = a_k s_k^{\sigma-1} \frac{1}{\sum_{l=1}^K \alpha_l s_l^\sigma} \frac{Y}{N},$$

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

E.2 Solution and Estimation with Aggregate Uncertainty

Our model features heterogeneity in worker types and two aggregate sources of heterogeneity, z and δ . 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 (18), which

Figure A7: Labor Productivity: Empirical Impulse Responses to a Separation Shock



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 (15) 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 A9).

depend on the job finding and separation rates, describe the distributional movements over time.

Therefore, there is no need to revert to iterative procedures, as the law of motion for the distribution is known a priori. We solve the model using first order perturbation around its stationary steady state (i.e., including the employment distribution). The large number of state variables (the two aggregate shocks, the distribution of employment 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 1600. 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.

E.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 4.8 we use data on average labor productivity and the share of newly hired workers in employment to estimate the time path of the two

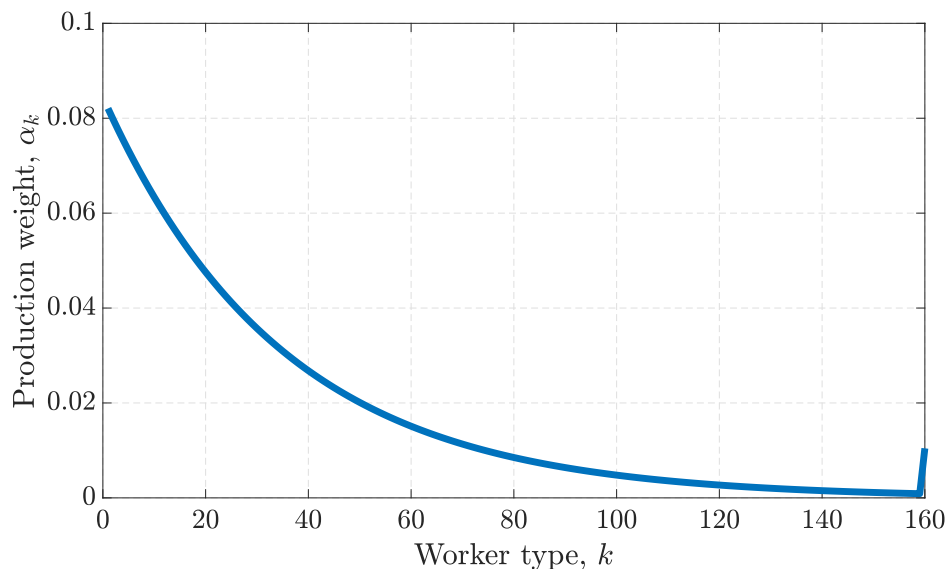
aggregate shocks consistent with these two time series and our parameterization. The model structure then implies a particular time path for all model variables. We use this property in Section 4.8 to calculate the contribution of congestion unemployment to the variation in observed unemployment fluctuations. Figure A15 shows the time paths of other labor market variables implied by our estimation.

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

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

Figure A8 visualizes the calibrated values of the relative productivities. Their pattern mimics that of employment shares. Relatively abundant types, such as worker type $k = 1$, would be characterized by a lower marginal product unless its abundance is offset by a higher relative productivity weight α_1 . The spike at $k = K$ is due to the fact that this type is an absorbing state and therefore employment in this type is somewhat higher than in $k = K - 1$.

Figure A8: Relative Worker Productivities in the Congestion Model



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

G 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.

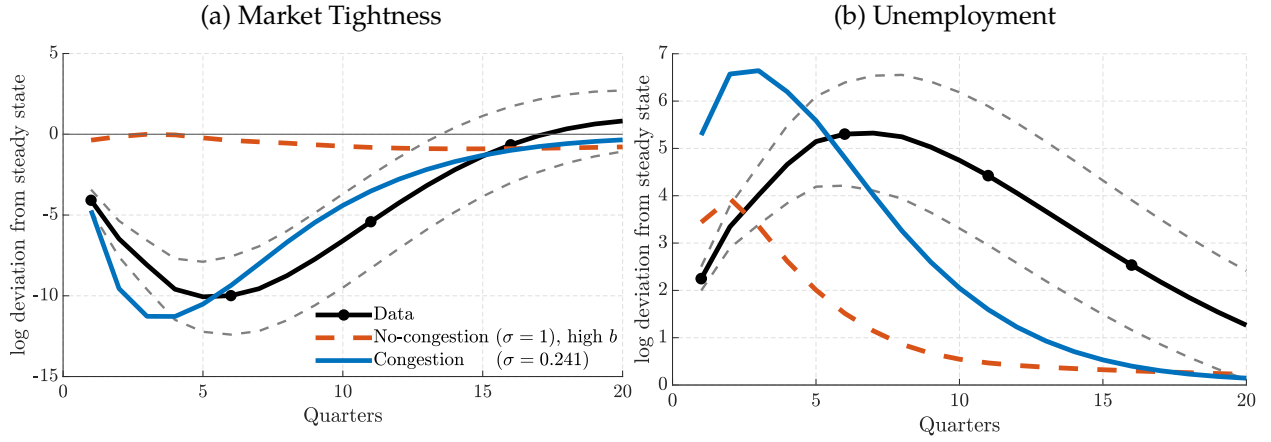
Results are presented in Table A2. While the model without separation shocks matches—by construction—the volatility of labor market tightness, it fails on the cyclical-ity of UE flows, for the same reasons as discussed in Section 2.1: separation shocks are necessary to match the countercyclical nature 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 countercyclical-ity 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 K above.

Figure A9 shows the empirical response of labor market tightness to a separation shock, with that of the model without congestion but with a low fundamental surplus and separation shocks. As in the standard model without congestion, there is essentially no response of labor market tightness to a separation shock. This key result does not change with a low fundamental surplus.

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

$$r + \delta = \frac{(z - b)(1 - \phi)q(\theta)}{\kappa} - \phi f(\theta), \quad (\text{A7})$$

Figure A9: Impulse Responses to a Separation Shock: No-Congestion, Low-Surplus Model



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

where r is the interest rate such that $\beta = 1/(1+r)$. Taking z as given and Implicitly differentiating Equation (A7) with respect to δ and θ gives

$$\begin{aligned} d\delta &= \frac{(z-b)(1-\phi)q'(\theta)}{\kappa} d\theta - \phi f'(\theta) d\theta \\ &= -[\mu(r+\delta) + \phi f(\theta)] \frac{d\theta}{\theta}. \end{aligned} \quad (\text{A8})$$

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

$$\epsilon_{\theta,\delta} = \frac{d\theta/\theta}{d\delta/\delta} = -\frac{\delta}{\mu(r+\delta) + \phi f(\theta)} = -\Upsilon^{Nash} \frac{\delta}{r+\delta + \phi f(\theta)}, \quad (\text{A9})$$

where $\Upsilon^{Nash} = \frac{r+\delta+\phi f(\theta)}{\mu(r+\delta)+\phi f(\theta)}$ is the scaling factor, which multiplies the fundamental surplus, derived in Ljungqvist and Sargent (2017). As discussed in Ljungqvist and Sargent (2017), reasonable calibrations of the standard search and matching model results in $\Upsilon^{Nash} \approx 1$. Moreover, these calibrations also result in the denominator in (A9) being roughly equal to one half. In conclusion, the standard model features labor market tightness which is largely insensitive to separation shocks, with an elasticity of around -2δ . Moreover, this elasticity is *independent* of the fundamental surplus. This is precisely the reason why even a calibration with a low fundamental surplus cannot replicate the empirical response of labor market tightness to separation shocks.

Table A2: Business Cycle Properties: No-Congestion, Low-Surplus Model

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
<i>Panel A: Low fundamental surplus model: no δ shocks</i>							
Standard deviation	0.010	0.064	0	0.052	0.199	0.230	0.049
Autocorrelation	0.706	0.706	0	0.844	0.596	0.706	0.311
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.999	1					
δ	–	–	–				
<i>u</i>	–0.647	–0.648	–	1			
<i>v</i>	0.980	0.981	–	–0.486	1		
θ	0.999	1.000	–	–0.648	0.981	1	
<i>UE/E</i>	0.476	0.476	–	–0.270	0.477	0.476	1
<i>Panel B: Low fundamental surplus model: with δ shocks</i>							
Standard deviation	0.010	0.064	0.082	0.090	0.177	0.227	0.068
Autocorrelation	0.691	0.689	0.560	0.825	0.558	0.689	0.623
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.999	1					
δ	–0.413	–0.430	1				
<i>u</i>	–0.674	–0.684	0.699	1			
<i>v</i>	0.933	0.929	–0.197	–0.368	1		
θ	0.999	1.000	–0.430	–0.684	0.929	1	
<i>UE/E</i>	0.005	–0.001	0.266	0.455	0.229	–0.001	1

Notes: *ALP*, *f*, δ , *u*, θ , *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 1600.

H Alternative Calibration: Matching EU Flows

The baseline model overpredicts the volatility of employment-to-unemployment (EU) flows in the data and underpredicts their persistence. The reason is that it is calibrated to match the volatility and persistence of UE flows as a share of employment, which drive our congestion mechanism.

This appendix provides an alternative calibration in which we match the time series properties of EU flows directly. This target provides a much tighter fit of the separation rate process. A cost of this alternative calibration is that the model-implied UE flows as a share of employment are less volatile than in the data, which weakens our congestion mechanism. However, even this model still manages to account for a large share of labor market volatility (e.g. 83 percent of unemployment volatility) while, at the same time, featuring a negative Beveridge curve (albeit weaker compared to the baseline model).

H.1 Measured and Implied Employment to Unemployment Flows

Appendix A describes the measurement of empirical time series of EU and UE flows in detail. In our baseline measurement, we ignore flows in and out of the labor force and adopt a two-state setting. One consequence is that the following relationships—which hold in the model at all times—do not hold for the empirical measures:

$$u_{t+1} = (1 - f_t)u_t + \delta_{t+1}(1 - u_t)$$

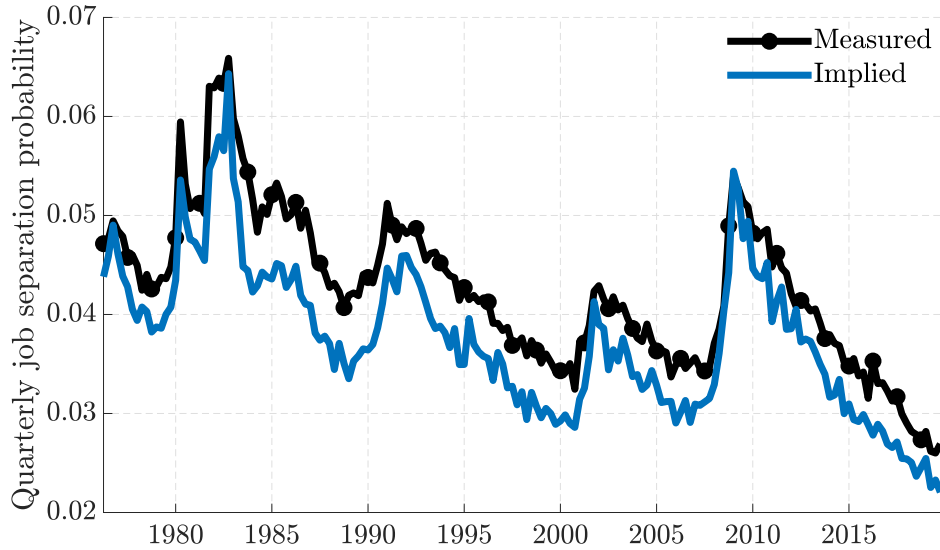
However, it is possible to compute a measure of *EU flows consistent with the two-state law of motion of unemployment* given by the above equation and the measured unemployment and job finding rates in the data.² Specifically, we compute the implied δ process as

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

Figure A10 shows the measured and implied time series for δ . The implied time series—consistent with the structure of our model—is more volatile and less persistent compared to the measured one. In particular, the standard deviation of the cyclical component of the measured δ is 0.067 while it is 0.084 for the implied one. The autocorrelation coefficient is 0.77 for the measured δ and it is 0.75 for the implied δ . Therefore, the implied measure

²This 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 A10: Separation Rate: Measured and Implied



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

of δ is closer in its time series properties to the δ process our model estimates on the basis of gross flows (UE).

H.2 Recalibrating the Model Using Information on EU Flows

From the above discussion it is clear that matching both the EU flows and UE flows as a share of employment is simply not possible by the model. The baseline model focuses on matching UE flows as a share of employment because they are central to the congestion mechanism. We now additionally quantify the performance of our model if it were, instead, calibrated to match the properties of the (implied) δ process described above.

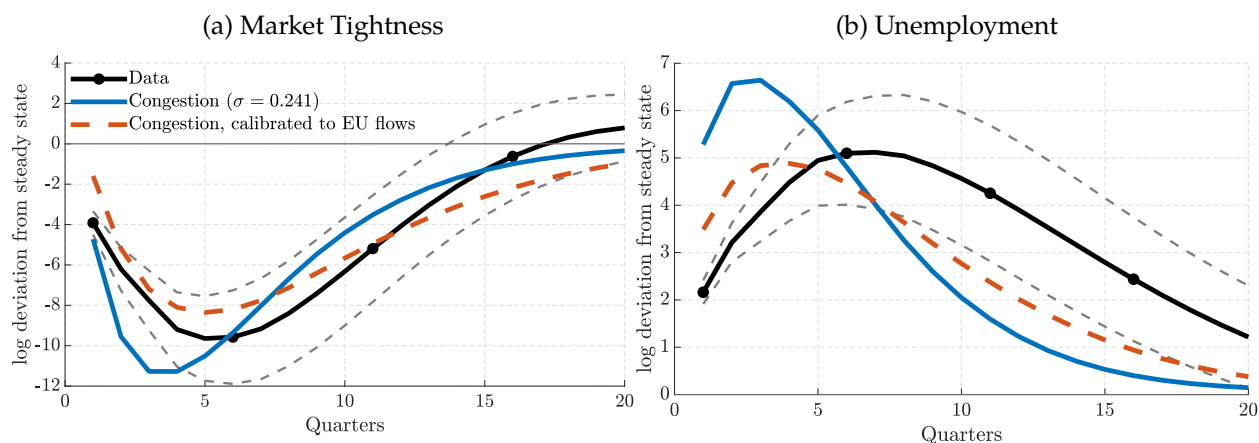
In particular, we recalibrate the process for δ (persistence ρ_δ and volatility σ_δ) and the degree of substitutability in production (σ) such that the model jointly matches the volatility and persistence of δ^{imp} (denoting the implied measure of δ based on the unemployment law of motion) and the empirical impulse response of θ with respect to δ^{imp} , as is the case in the baseline.³

Figure A11 shows the impulse responses of unemployment and labor market tightness to δ in the data, the baseline model and the alternative calibration. Figure A12 plots the IRFs of δ on itself.

Table A3 reports the business cycle statistics in the data, the baseline model calibrated to UE flows as a share of employment and the alternative calibration in which the base-

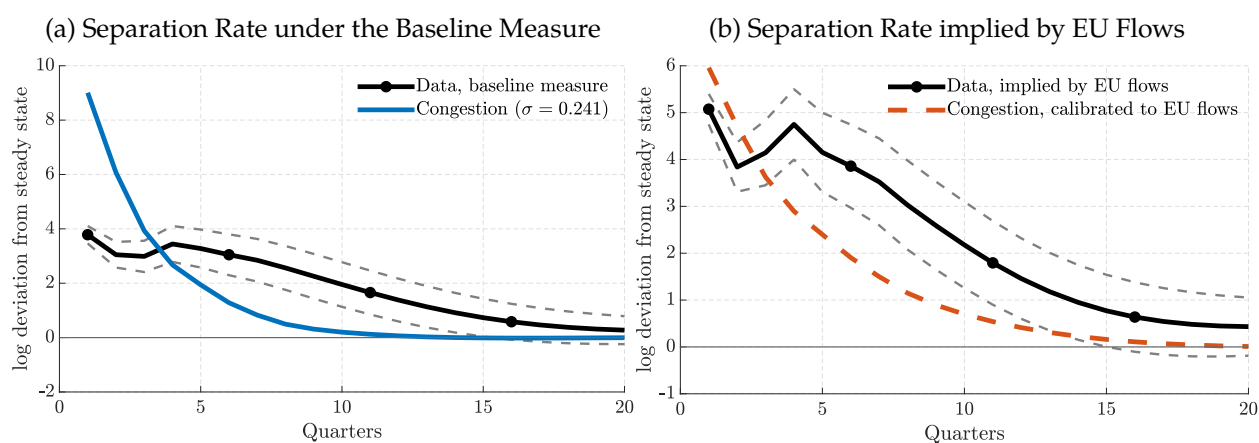
³In doing so, we require that the RMSE of the impulse response does not exceed the one in the baseline model.

Figure A11: 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.

Figure A12: Impulse Responses of Separation Rate to a Separation Shock: Baseline and Alternative Calibration of Separation Rate Process to Match EU Flows



Notes: The figure plots the impulse responses of the separation rate to a separation rate shock in the data and the corresponding model calibrations: the baseline calibration and the alternative specification that is calibrated to match the business cycle patterns of EU flows.

line model is parameterized to δ . In the alternative calibration, the volatility of UE/E is about 20 percent lower than in the data. Therefore, our congestion mechanism is weaker. Counterfactually weaker congestion reduces the volatility of job finding rates and unemployment, to about 83 percent that of the data. But it also worsens the comovement statistics; most importantly, the Beveridge curve is now weaker at -0.56 .

Table A3: Business Cycle Properties of the Congestion Model: Baseline and Alternative Calibration

	ALP	f	δ^{imp}	u	v	θ	UE/E	p_1
<i>Panel A: Congestion Model - Calibrating to UE/E</i>								
Standard deviation	0.010	0.059	0.122	0.121	0.102	0.207	0.067	0.055
Autocorrelation	0.688	0.897	0.530	0.836	0.857	0.897	0.742	0.771
Correlation matrix								
ALP	1							
f	0.443	1						
δ	-0.410	-0.509	1					
u	-0.463	-0.924	0.743	1				
v	0.348	0.922	-0.157	-0.716	1			
θ	0.443	0.996	-0.514	-0.940	0.909	1		
UE/E	-0.337	-0.930	0.392	0.865	-0.876	-0.940	1	
p_1	0.490	0.952	-0.431	-0.862	0.900	0.949	-0.973	1
<i>Panel B: Congestion Model - Calibrating to δ</i>								
Standard deviation	0.010	0.041	0.084	0.086	0.077	0.144	0.052	0.054
Autocorrelation	0.688	0.920	0.619	0.866	0.821	0.920	0.761	0.785
Correlation matrix								
ALP	1							
f	0.359	1						
δ	-0.425	-0.431	1					
u	-0.424	-0.884	0.766	1				
v	0.197	0.875	0.045	-0.560	1			
θ	0.358	0.997	-0.434	-0.898	0.866	1		
UE/E	-0.309	-0.906	0.518	0.887	-0.717	-0.914	1	
p_1	0.468	0.916	-0.550	-0.878	0.726	0.913	-0.970	1

Notes: ALP , f , δ , u , θ , UE/E and p_1 indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness, share of new hires in employment and the marginal product of labor of new hires. Panel A reports values from the baseline model; Panel B reports the values for alternative calibration in which EU flows are targeted (rather than UE flows). All variables have been logged and the empirical cyclical components have been extracted using the HP-filter with a smoothing parameter of 1600.

Importantly, even this congestion model with counterfactually little congestion still outperforms the standard model with separation shocks. Table A5 shows that the standard model calibrated to match the UE/E fluctuations features separation shocks which are *more volatile* than the data. Nevertheless, the volatility of unemployment is only at 2/3 of the data and the Beveridge curve is strongly *positively* sloped (correlation of unemployment and vacancies of 0.97).

I Deriving the Iso-congestion Curve

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

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

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

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

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

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

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

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

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

elasticity of z :

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

Fourth, using this identity and the fact that $\varepsilon_{p_k^{nc}, n_k} = 0$, we derive our desired elasticity of marginal product with respect to worker mass:

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

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

$$\varepsilon_{p_k, n_k} = (1-x)(\sigma - 1) \left(1 - \frac{\alpha_k^c n_k^\sigma}{\sum_k \alpha_k^c n_k^\sigma} \right). \quad (\text{A10})$$

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

$$\varepsilon_{p_k, n_k} = (1-x)(\sigma - 1)(1 - s_k). \quad (\text{A11})$$

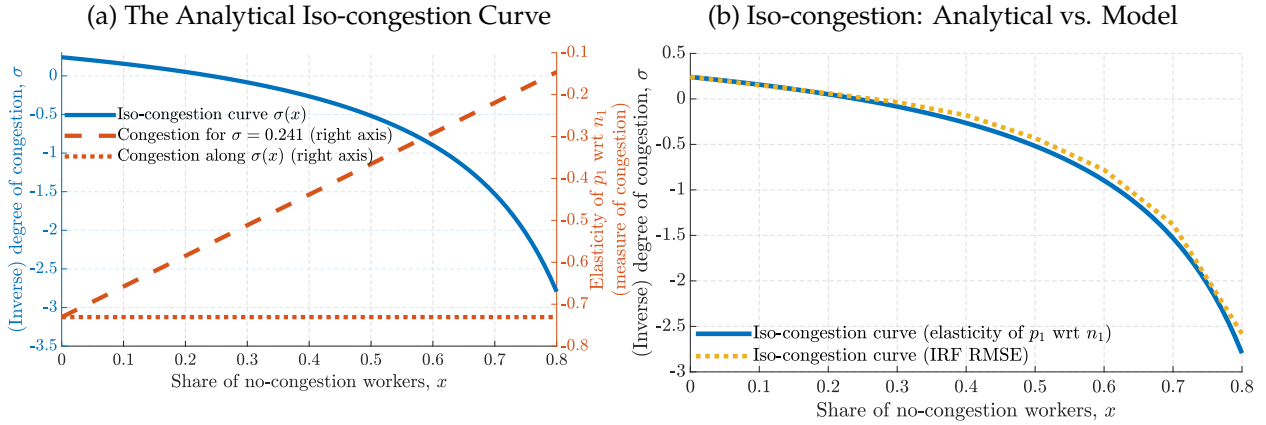
To trace out the iso-congestion curve for $k = 1$, we solve for σ as a function of x given a level of elasticity $\bar{\varepsilon}_{p_1, n_1}$.

$$\sigma(x) = 1 + \frac{\bar{\varepsilon}_{p_1, n_1}}{(1-x)(1-s_1)}. \quad (\text{A12})$$

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

Figure A13 Panel (a) plots the iso-congestion curve derived in Equation (A12) starting from our baseline calibration of $x = 0$ and $\sigma = 0.241$. The figure makes clear that, as there

Figure A13: Iso-congestion Curves



Notes: Panel (a) plots the analytical iso-congestion curve as a function the share of no-congestion workers in production, x . It also includes the level of congestion as a function of x , as well as the constant level maintained along the iso-congestion curve. Panel (b) compares the analytical iso-congestion curve to the one we obtain solving our dynamic congestion model by matching the IRF of labor market tightness to the separation rate shock in Figure 9 Panel (a).

is more weight on no-congestion workers in final good production, σ needs to be adjusted downward to maintain the same level of congestion as in our baseline calibration. In fact, if $\sigma = 0.241$ is held constant, higher levels of x lead to smaller congestion in production.

Panel (b) superimposes the iso-congestion curve we present in the main text based on the solution to the full dynamic model and on matching the IRF of labor market tightness to the separation rate shock in Figure 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.

J 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:⁴

$$c(UE_t) = c_1 \cdot \left[\left(\frac{UE_t}{UE_{ss}} \right)^{c_2} - 1 \right]. \quad (\text{A13})$$

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

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

$$\frac{\kappa}{q_t} + c(UE_{t+1}) = E_t [\beta(1 - \delta_{t+1})J_{t+1}]. \quad (\text{A14})$$

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

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

As with the production-function based congestion parameter σ , we now set c_2 such that the model minimizes the RMSE of the response of labor market tightness to separation shocks. Figure A14 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 A4. 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.)

⁴Pissarides (2009); Silva and Toledo (2013) add a fixed costs of hiring, but it is not increasing in the amount of hires.

Figure A14: Impulse Responses to a Separation Shock: Convex Hiring Cost Model



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

Table A4: Business Cycle Properties: Convex Hiring Cost Model

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

Notes: ALP , f , δ , u , θ , 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, 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 1600.

K 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 A5). In the main text, Table 2 documents that the congestion model succeeds in replicating these patterns. In contrast, Table A5 documents the failures of the standard model.

In particular, the standard DMP model without separation shocks cannot match the cyclicity 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.⁵

In Appendix G, 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.

⁵In 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.

Table A5: Business Cycle Properties of the No-Congestion Model

	<i>ALP</i>	<i>f</i>	δ	<i>u</i>	<i>v</i>	θ	<i>UE/E</i>
<i>Panel A: Standard model without separation shocks</i>							
Standard deviation	0.010	0.004	0	0.003	0.013	0.015	0.003
Autocorrelation	0.704	0.704	0	0.843	0.592	0.704	0.306
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	1.000	1					
δ	–	–	1				
<i>u</i>	–0.643	–0.643	–	1			
<i>v</i>	0.980	0.980	–	–0.481	1		
θ	1.000	1.000	–	–0.643	0.980	1	
<i>UE/E</i>	0.476	0.476	–	–0.272	0.476	0.476	1
<i>Panel B: Standard model with separation shocks</i>							
Standard deviation	0.010	0.005	0.088	0.068	0.058	0.017	0.067
Autocorrelation	0.688	0.647	0.499	0.736	0.751	0.647	0.74
Correlation matrix							
<i>ALP</i>	1						
<i>f</i>	0.975	1					
δ	–0.441	–0.627	1				
<i>u</i>	–0.508	–0.665	0.916	1			
<i>v</i>	–0.306	–0.482	0.888	0.974	1		
θ	0.975	1.000	–0.627	–0.665	–0.482	1	
<i>UE/E</i>	–0.348	–0.402	0.413	0.739	0.747	–0.402	1

Notes: *ALP*, *f*, δ , *u*, θ and *UE/E* indicate, respectively, average labor productivity, the job finding rate, separation rate, unemployment rate, labor market tightness and the share of new hires in employment. 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 1600.

L Historical Decomposition: Additional Material

The main text shows how congestion-only unemployment contributed to the evolution of overall unemployment. In this Appendix, we provide the same exercise also for TFP- and separation-driven unemployment. The estimated time paths of key labor market variables are presented in Figure A15.

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 z_t shocks only, u^z , which would arise in the TFP-shock-only models such as in Shimer (2005); Hall (2005b); Hagedorn and Manovskii (2008) and generated by separation shocks only, u^δ . The corresponding equations that characterize these counterfactuals are, for u^z ,

$$\begin{aligned} u_{t+1}^z &= (1 - f(\theta_t^z))u_t^z + \bar{\delta}(1 - u_t^z), \quad \kappa = q(\theta_t^z)\beta\mathbb{E}_t(1 - \bar{\delta})S_{1,t}^z \\ S_{k,t}^z &= z_t - b + \beta\mathbb{E}_t(1 - \bar{\delta})S_{k+1,t+1}^z - \beta\mathbb{E}_t(1 - \bar{\delta})f(\theta_t^z)\phi S_{1,t+1}^z \text{ for all } k, \end{aligned} \quad (\text{A15})$$

and, respectively, for u^δ ,

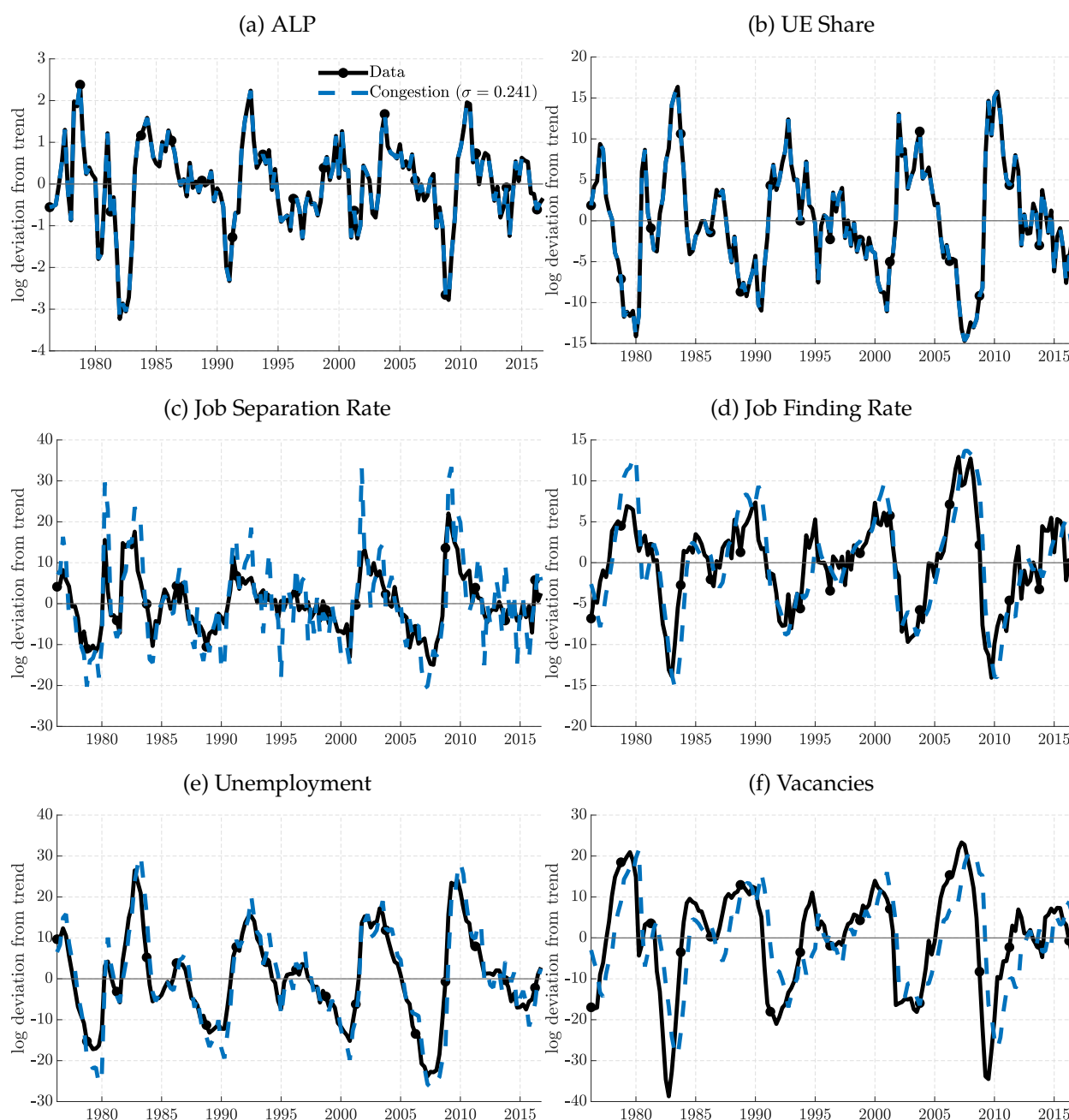
$$\begin{aligned} u_{t+1}^\delta &= (1 - f(\theta_t^\delta))u_t^\delta + \delta_{t+1}(1 - u_t^\delta), \quad \kappa = q(\theta_t^\delta)\beta\mathbb{E}_t(1 - \delta_{t+1})S_{1,t}^\delta \\ S_{k,t}^\delta &= \bar{z} - b + \beta\mathbb{E}_t(1 - \delta_{t+1})S_{k+1,t+1}^\delta - \beta\mathbb{E}_t(1 - \delta_{t+1})f(\theta_t^\delta)\phi S_{1,t+1}^\delta \text{ for all } k. \end{aligned} \quad (\text{A16})$$

Figure A16 plots the associated time series of these counterfactual unemployment rates together with actual unemployment. Table A6 provides a set of business cycle statistics related to overall unemployment and the three counterfactuals.

Volatility. Table A6 quantifies the role of congestion-driven unemployment in U.S. 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.⁶

⁶As discussed in Section 4.4, our model matches UE flows by estimating a somewhat more volatile separation rate process. In Table A6, 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.

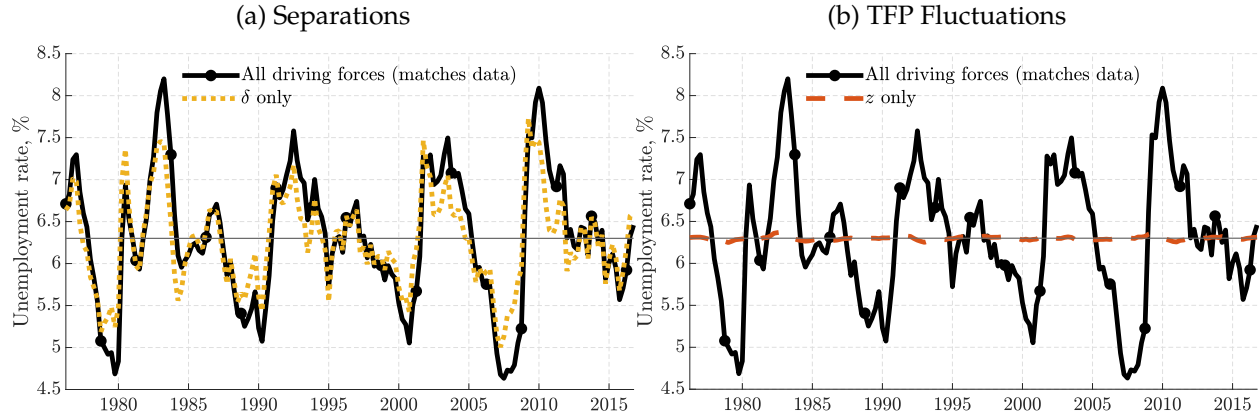
Figure A15: Time Paths of Labor Market Variables



Notes: The figure plots the estimated time paths of labor market variables using the Kalman Filter. Time series are logged and HP-filtered using a smoothing parameter of 1600.

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

Figure A16: Unemployment Components



Notes: The figure plots actual, and counterfactual unemployment rates u^z and u^δ estimated using data on the cyclical components of average labor productivity and new hires as a share of employment. The counterfactual unemployment time series are based on Equations (A15) and (A16).

Table A6: Historical Decomposition of Unemployment: Model and Counterfactuals

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

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

mechanisms laid out in Section 4.7.