Worker heterogeneity, selection, and employment dynamics in the face of aggregate demand and pandemic shocks*

Federico Ravenna† and Carl E. Walsh‡

July 10, 2020

Abstract

In a new Keynesian model with random search in the labor market, endogenous selection among heterogeneous workers amplifies fluctuations in unemployment and results in excess unemployment volatility relative to the efficient allocation. Recessions disproportionately affect low-productivity workers, whose unemployment spells are inefficiently frequent and long. We consider a COVID-recession resulting from a negative demand shock and a surge in exogenous separations. High-productivity workers benefit if separations in a pandemic take the form of temporary layoffs, but this is not true for low-productivity workers. The unemployment consequences are especially severe when nominal interest rates are close to the effective lower bound.

Keywords: Unemployment, heterogeneity, selection, COVID-19, ZLB
JEL classification: E24, E32, E52.

Since 2007 the U.S. has experienced two severe contractions driven by demand shocks – the Great Recession and the COVID recession – set off by the financial crisis in the first case and by labor market disruptions and demand shortfalls associated with economic shutdowns in the second. We show that these types of recessions have a disproportionate impact on those workers who, on average, have more frequent and longer spells of unemployment. Employing a new Keynesian model with search and matching frictions in the labor market, combined with a simple model of worker heterogeneity, we show how heterogeneity results in procyclical

*This paper was prepared for the National Bank of Ukraine Annual Research Conference on "Labor Market and Monetary Policy", 28–29 May 2020. We are very grateful for comments from conference participants, and from Bart Hobijn, Renato Faccini, Yuriy Gorodnichenko, Oleksiy Kryvtsov, Karl Walentin, and Kenneth West.

†Head of Research, Danmarks Nationalbank, University of Copenhagen, HEC Montreal and CEPR. Email: fera@nationalbanken.dk.

‡Distinguished Professor of Economics, University of California, Santa Cruz. Email: walshc@ucsc.edu.

fluctuations in the average productivity among the pool of job seekers, and how this has important implications for the response of employment, hours, and wages in the face of a severe contractionary shocks. These effects arise because of heterogeneity-driven selection in both job separations and hiring; firm-worker matches that end are not simply chosen randomly from among existing matches and firms selectively screen job seekers before making hires. We show how selection matters even in a COVID-19 pandemic scenario in which there is a surge in mass layoffs that initially affect all workers non-selectively. Modeling the pandemic as due to both a negative shock to the demand for goods and services and to a surge in exogenous separations, the resulting fall in match surplus results in additional endogenous separations, amplifying the resulting rise in unemployment among those workers with worse average lifetime labor market outcomes.

This result obtains even when parameterizing the model to account for the large rise in the share of temporary layoffs observed in the COVID recession. The unemployment consequences of a negative demand shock are especially severe when monetary policy is constrained by the effective lower bound on nominal interest rates. Finally, relative to a model that assumes homogeneous labor, heterogeneity among workers generates fluctuations in unemployment that are both large and inefficient.

The paper makes five contributions. First, it shows how selection arising from labor heterogeneity affects macroeconomic dynamics in response to aggregate demand shocks and to shocks that generate mass layoffs. We find that even when these shocks generate the same fall in aggregate output, they generate quite different behavior of unemployment and hours. Second, those workers who experience more frequent and longer average unemployment spells are adversely affected by recessionary shocks and, relative to the social planner's allocation, these workers experience inefficiently high job separation rates and inefficiently low job finding rates during recessions. Third, in a COVID-pandemic scenario, selection amplifies the resulting rise in unemployment but has differential effects on wages of high- and low-productivity workers. Fourth, the impact of a COVID-recession on unemployment and deflation is further amplified when the monetary policy is constrained by the effective lower bound on nominal interest rates. Finally, if layoffs in a pandemic are predominately temporary, the recession still triggers a rise in endogenous separations, and while unemployment among high-productivity workers returns quickly to its steady-state value, that of low-productivity workers and the overall unemployment rate only slowly returns to normal levels.

The separation rate of workers from employment, as well as the share of laid-off workers within the unemployed, has been documented to be countercyclical in postwar United States (Fujita and Ramey 2009). Importantly, workers experiencing layoffs in a recession are not randomly selected from the pool of employed workers; workers with certain characteristics (young, low-schooling, etc.) experience higher increases in joblessness during a downturn. Several authors (see Grigsby 2020; Baley 2020) find that these workers also differ in unobservable characteristics; ceteris paribus, they have lower productivity. Grigsby 2020 estimates that selectivity in separations and hiring during the Great Recession lead to the efficiency of production workers rising by 50%, and to a 10% rise in the aggregate mean human capital of employed workers economy-wide.

The heterogeneity we focus on arises from *ex ante* unobservable differences among workers. Workers differ along many dimensions, and some, such as educational level, specific job skills or experience, age, and gender may be easily observable. Other differences are difficult for prospective employers to observe, and Mincer-wage regressions that condition on observable characteristics of workers exhibit large unexplained residual variation in wages across workers (see Lemieux (2006), and Hornstein et al. (2011)). Other aspects of labor market outcomes are also difficult to explain based on observable worker characteristics. For example, Dickens and Triest (2012) estimate a multinormal logit model of involuntary separation transition probabilities using the 2008 wave of the SIPP. Controlling for age, education, race, and gender, their estimated equation has an R-squared of 0.129, suggesting heterogeneity of worker experiences within groups classified based on standard observable characteristics is important. It is this *ex-ante unobservable heterogeneity* that is the focus of our model. Thus, the effects we emphasize would still operate if labor markets were segmented by observable characteristics. When workers are heterogeneous, selectivity in separations and hiring has a large impact on the aggregate economy, resulting in an inefficient allocation and excess volatility of unemployment.

Our framework builds on the model of Ravenna and Walsh (2012b) which combines the basic structure of a new Keynesian model with a frictional labor market and worker-specific productivity heterogeneity.[1] We assume workers are of two types: high average productivity and low average productivity. While an unemployed worker's type is unobserved ex ante, we assume a firm can observe the productivity type of its existing employees. A firm that is hiring engages in a process of interviewing, or screening, during which the firm is able to observe the productivity of a job applicant. The aggregate separation rate consists of exogenous and endogenous components. Firms employ an (optimal) cutoff productivity strategy; any job applicant whose productivity exceeds the cutoff is hired; any existing worker whose productivity is below the cutoff is fired. This cutoff productivity is endogenous, so the aggregate job separation rate is endogenous.

With separations generated through selection, a rise in separations generates a compositional effect on the pool of unemployed that reduces the expected productivity of a prospective job seeker. This affects the vacancy posting decision of firms and the job filling rate. Selection is relevant even in a recession associated with a surge in mass layoffs that initially affects all worker types. As the economy recovers, matches need to be re-established, and reestablishing matches involves selection; firms need to meet and screen new workers of ex-ante uncertain ability.

We find that a COVID-induced recession driven by mass layoffs and a negative demand shock reduces match surplus, leading to an increase in endogenous separations that amplify the unemployment effect of the recession. Because endogenous separations are done selectively, the productivity of the average employed worker rises relative to that of average unemployed worker. An rise in separations among all workers ends up affecting low-productivity workers differentially. They see the largest rise in joblessness, both in absolute and in percentage terms, as the selection that characterizes the increase in endogenous separations strongly biases the adverse impact of the COVID shock towards the low-productivity group of workers.

The U.S. data for the first three months of the COVID-recession suggests on balance the pandemic shocks have been deflationary, a result consistent with our model. We find that a binding effective lower bound on the nominal interest rate that limits the reaction of monetary policy leads, when paired with selection in the adjustment of employment, to a short lived but

[1] Our model is as close as possible to a baseline new Keynesian model of the business cycle with search and matching in the labor market. We assume the only propagation mechanism is the endogenous selection, and the only inefficiencies in the allocation stems from selection and price stickiness.

large additional increase in unemployment, significantly amplifying the impact of the COVID shock on unemployment by 50% at peak, and leading to a further fall in inflation.

We show the decisions to screen out workers in the interview process and to separate selectively among employed workers creates an externality due to the impact of these decisions on the composition of the pool of unemployed workers. This distortion is the result of an externality similar but separate from the congestion externality well known to exist in model of the business cycle with search and matching frictions in the labor market. It results in excess unemployment in a recession, and in a relatively larger weight of the total increase in unemployment being borne by low-productivity workers, who have a lower expected wages and longer unemployment durations. This creates a distortion, making the competitive equilibrium inefficient even when the standard distortions present in new Keynesian models with search and matching frictions are eliminated.

Selection also results in a marked disconnect between labor productivity, unemployment and wages. Selection makes wages endogenously sticky - despite our assumption that wages are Nash bargained in every period. Wage growth can be zero even while unemployment keeps decreasing for several periods during a recovery. The wage-gap is closed far earlier than the unemployment gap during the recovery from a downturn.

Our paper is related to three areas of the literature. First, worker and match heterogeneity play a key role in several models in the search and matching literature and in models with job-to-job transitions (e.g., Guerrieri (2007), Nagypal (2007), Nagypal and Mortensen (2007), Hall and Schulhofer-Wohl (2018)). Bils et al. (2012) study the implications of skill heterogeneity for wages and labor market flows over the business cycle, but they assume segmented labor markets and only consider productivity shocks. We focus on heterogeneity based on unobservable that preclude full market segmentation. Pries and Rogerson (2005) allow for persistent job-specific productivity variation, and firms screen workers based on limited information on their productivity. They assume the average productivity of unemployed workers is not state-dependent, and the authors focus on steady state results rather than on the dynamics of labor market variables over the business cycle.

In a model with heterogeneous skills and exogenous separation rates, Pries (2008) shows that the composition effect has a large impact on the cyclical value of vacancies and thus on the behavior of employment flows. Ahn and Hamilton (2019) emphasize *unobserved* differences across workers in (exogenous) unemployment exit probabilities, consistent with the idea that heterogeneity among the workers flowing into unemployment can account for differences in future outflow rates. This heterogeneity hypothesis (see Davis 1996 and Baker 1992) is central to our approach. Ravn and Sterk (2017) also develop a model with two worker types, but they focus on differences in search efficiency rather than productivity differences, and they assume separation probabilities are the same for both types – only job finding rates differ. We allow both separation rates and job finding rates to vary endogenously and to differ across worker types. While the framework we propose is closely related to this previous work on labor heterogeneity in a search and matching environment, we provide a model with nominal rigidities that allows aggregate demand and the role of monetary policy to be analyzed.

Second, our modeling framework is related to several contribution in the literature on labor and nominal frictions. We include nominal rigidities in a model with unemployment. Earlier contributors to this area include, among others, Walsh (2003), Trigari (2009), Blanchard and Galí (2007, 2010), Gertler et al. (2008), Gertler and Trigari (2009), Ravenna and Walsh (2008, 2011, 2012a), Galí (2011), and Christiano et al. (2016). However, these contributions with the exception of that of Walsh, assume an exogenous separation rate, and all assume homogenous workers. We show how layoffs in the competitive equilibrium can be inefficient, a result that is consistent with that of Berger et al. (2019), who argue for monetary policy to target the layoff rate in a model with countercyclical layoffs.

Finally, a growing number of papers have modelled the macroeconomic implications of COVID-19. Guerrieri et al. (2020) focus on sectorial heterogeneity in a model with two sectors to show how job destruction in one sector can create a demand-driven recession in the other sector. However, the evidence in Kahn et al. (2020) suggests employment loss in April 2020 as measured by unemployment insurance claims were common across U.S. industries and occupations, whether the industry was considered essential or work-from-home capable. Kapicke and Rupert (2020) focus on employment adjustments caused by a pandemic within a search and matching framework. They distinguish between workers by health status and investigate the evolution of health and employment status in a real model. Gregory et al. (2020) study the effects of COVID-19 in a search model with worker and sector heterogeneity. They assume workers' type distribution may differ across sectors, but transition probabilities between states are partly exogenous, depending on the workers type. Our focus is on how selection affects those probabilities through the impact of shocks on optimal labor market choices. They also assume a constant discount rate and abstract from nominal rigidities, while we incorporate endogenous variation in the discount rate, a factor emphasized by Hall (2017) and found by Leduc and Liu (2020) to be important in explaining labor market fluctuations. Consistent with our results and those of Hall (2015) and Ravn and Sterk (2017), the low job-finding rates of some workers plays a crucial role in the behavior of unemployment during recoveries.[2]

The remainder of this paper is organized as follows. Our model is presented in section 1. Section 2 discusses the distortions present in the model that cause the competitive market equilibrium to differ from the social planner's allocation. We show that both the inflow into unemployment and the outflow into employment for low-efficiency workers are inefficiently high. The role of productivity heterogeneity and the composition effect is investigated in a calibrated version of the model in sections 3 and 4 where we discuss model dynamics in the response to a negative demand shock and a positive shock to exogenous separations respectively. Section 5 combines both these shocks to investigate a pandemic scenario. We think of this experiment as a way of investigating an economy-wide "shelter-in-place" that produces a jump in unemployment via mass layoffs that affects all workers equally while the economy also experiences a negative demand shock. Section 6 repeats the experiment when the ZLB on nominal interest rates restricts the ability of monetary policy to offset fully the negative demand shock. In section 7 we generalize the model to allow for temporary layoffs. We show temporary layoffs speed the employment recovery for high-productivity workers but have little impact on the recovery for low-productivity workers. Conclusions are summarized in the final section.

1 The model of productivity heterogeneity

The model consists of households, wholesale and retail firms, and a monetary policy authority. The representative household purchases consumption goods, holds bonds, and supplies labor to

[2] In analysing recovery after the Great Recession, Hall (2015) concludes that "The return to normal has been slower than in previous postrecession episodes because the crisis shifted the composition of job seekers toward those with low job-finding rates and low exit rates from unemployment." (p. 121)

wholesale firms. Wholesale firms hire labor and produce a homogeneous good that is sold in a competitive market to retail firms. The labor market is characterized by search and matching frictions. Retail firms transform the wholesale good into differentiated final goods which are sold to households for consumption and to wholesale firms to use in posting job vacancies. The key ingredient that differentiates the model from a standard NK plus search and matching model is our assumption that an individual worker can be one of two types. These types differ in terms of their average productivity as described below, and we refer to workers as being either of type l, for low average efficiency, or type h, for high average productivity. We focus our discussion on the specification of the labor market, as other aspects of the model reflect standard aspects of new Keynesian models. For more details, see Ravenna and Walsh (2012b).

1.1 Labor flows, wages and vacancies

The household consists of a continuum of workers; each is either employed or searching for a job. We assume complete consumption risk sharing among households. A fraction $\bar{\gamma}$ of workers are of low (l) average efficiency, while the remaining $1-\bar{\gamma}$ are of high (h) average efficiency.[3] The worker's efficiency-type, h or l, is permanently assigned. If L^j denotes the labor force of type j, $j = h, l$, $L^h + L^l = L = 1$. Let $N_t = N_t^l + N_t^h$ be total employment, where N^j be the number of type j workers who are employed, and let $\xi_t \equiv N_t^l / N_t$ be the fraction of employed workers of low efficiency.

The productivity of a type h worker equals ϕ^h. Regardless of whether employed or unemployed, each period all low-efficiency workers receive an idiosyncratic stochastic productivity shock so that the productivity of worker i of type l is $a_{i,t}^l \phi^l$. We assume $a_{i,t}^l$ is serially uncorrelated with support $(0\ 1]$ and cumulative distribution function $F(.)$. While productivity is stochastic for a low-efficiency worker, we assume it is constant, conditional on aggregate TFP, for a type h worker.[4]

A firm can observe the productivity of its existing employees. However, firms must interview

[3] Ahn and Hamilton (2019) show that the average duration of US unemployment can be matched if the labor force consists of just two types of workers who differ in their job finding probabilities, as will be the case endogenously in our model.

[4] This assumption is for simplicity as it will imply that endogenous separations and interviews that do not lead to hires only involve low skilled workers.In section 2, we discuss the inefficiency of the allocation when both types are treated symmetrically in experiencing idiosyncratic, stochastic fluctuations in productivity.

unemployed job applicants to determine a job seeker's efficiency type. The aggregate number of interviews per period is determined through random matching, and all job seekers have identical interview-finding probability, regardless of type. At the interview, the job applicant's productivity level and type is revealed. We assume the (nonstochastic) productivity of an h worker is sufficiently high to guarantee a positive surplus in all states. Thus, if the efficiency level is revealed to be h, the worker is hired and produces with probability equal to one. If an interview reveals the job seeker is a type l, firms employ an (optimal) cutoff strategy in which only type l workers with $a_{i,t}^{l} > \bar{a}_t$ will be hired; $\bar{a}_t$ will be endogenously determined and vary countercyclically. Consequently, conditional on being interviewed, the probably a type l worker is hired is $F(\bar{a}_t) < 1$. Absent any direct hiring and firing costs, $\bar{a}_t$ will also be the cutoff value for determining whether an existing employed low-efficiency worker generates a positive surplus and is retained.

At the start of each period, there is an exogenous separation probability ρ^x that affects all employed workers, regardless of efficiency level. Workers unmatched at the start of the period, or who do not survive the exogenous separation hazard, seek new interviews. There are

$$S_t = 1 - (1 - \rho^x)\, N_{t-1}$$

such job seekers. Let S^j be the number of type j workers who are seeking jobs and $S_t = S_t^h + S_t^l$ and denote the share of job seekers of quality l by $\gamma_t \equiv S_t^l / S_t$. We define the end-of-period number of unemployed workers as $U_t = 1 - N_t$.[5]

After exogenous separation occurs, all aggregate shocks realizations are observed and wholesale firms determine $\bar{a}_t$.[6] The time t idiosyncratic productivity shocks associated with employed low-efficiency workers and with job seekers who are interviewed are observed. With probability $\rho_t^n \equiv F(\bar{a}_t)$, a low-efficiency worker's productivity draw will be less than $\bar{a}_t$. An existing employee with productivity $a_{i,t}^{l} < \bar{a}_t$ becomes unemployed; an unemployed low-efficiency worker with $a_{i,t}^{l} < \bar{a}_t$ is not hired.

[5] The two measures of unemployment can differ as some job seekers find employment (and produce) during the period. In search models based on a monthly period of observation, it is more common to assume workers hired in period t do not produce until period $t+1$. In this case, the number of job seekers in period t plus the number of employed workers adds to the total work force. Because we base our model on a quarterly frequency, we allow for some workers seeking jobs to find jobs and produce within the same period.

[6] We show below that $\bar{a}_t$ is the same for all firms.

The number of vacancies posted by wholesale firms V_t, together with the number of job seekers, determines the number of interviews I_t via a standard CRS matching function:

$$I_t = \psi V_t^{1-a} S_t^a; \qquad 0 < \alpha < 1,\, \psi > 0. \tag{1}$$

A job seeker gets an interview with probability $k_t^w \equiv I_t/S_t = \psi \theta_t^{1-a}$, where $\theta_t \equiv V_t/S_t$. The job finding probability is identical to the interview rate for high-efficiency workers, while for low-efficiency workers it is lower, and equal to

$$k_t^{w,l} = (1 - \rho_t^n)\, k_t^w < k_t^w.$$

Because the probability a worker drawn from the pool of unemployed job seekers is low efficiency is γ_t, the overall job finding probability is

$$\gamma_t k_t^{w,l} + (1 - \gamma_t) k_t^w = (1 - \gamma_t \rho_t^n)\, k_t^w.$$

With heterogeneous workers, a job opening that would be filled and lead to production if a high-efficiency applicant is interviewed may go unfilled if a low-efficiency worker is interviewed. New hires H_t are given by the number of interviewees who are of high efficiency, all of whom are hired, plus the number of interviewees who are of low efficiency times the fraction of these with productivity levels that exceed $\bar{a}_t$:

$$H_t = (1 - \gamma_t) k_t^w S_t + (1 - \rho_t^n)\, \gamma_t k_t^w S_t = (1 - \gamma_t \rho_t^n)\, k_t^w S_t. \tag{2}$$

Fewer workers are hired than are interviewed: $H_t < k_t^w S_t$. Screening implies new hires depend on the endogenous average quality of the pool of unemployed workers as measured by γ_t and on the aggregate productivity and markup which we show below will affect $\bar{a}_t$ and therefore ρ_t^n.

Low-efficiency workers employed in existing matches that survived the exogenous separation hazard also receive a new productivity shock and are retained if and only if $a_{i,t}^l > \bar{a}_t$. Thus, actual employment in period t is equal to

$$\begin{aligned} N_t &= (1-\rho^x)\left[\left(1-\xi_{t-1}\right)+\xi_{t-1}(1-\rho_t^n)\right]N_{t-1}+H_t \\ &= (1-\rho^x)\left(1-\xi_{t-1}\rho_t^n\right)N_{t-1}+H_t \end{aligned}$$

The share of low efficiency employed workers evolves according to

$$\xi_t = (1-\rho_t^n)\left[\frac{(1-\rho^x)\xi_{t-1}N_{t-1}+\gamma_t k_t^w S_t}{N_t}\right]. \tag{3}$$

Job seekers at t who are of quality l equal the total number of low-efficiency workers minus the number of matches of quality l that survive the exogenous separation hazard. Hence,

$$\gamma_t = \frac{L^l-(1-\rho^x)\xi_{t-1}N_{t-1}}{S_t}. \tag{4}$$

The efficiency-weighted average productivity of both employed workers and the pool of job seekers will change over time because γ_t is endogenous and persistent, even though $a_{i,t}^l$ is *i.i.d.*. During recessions, the outflow from employment rises and the inflow into employment falls, resulting in an increase in the average productivity among those still employed and a fall in the average efficiency level of those who are unemployed.

Labor is used by wholesale firms to produce a homogenous output that is sold in a competitive market at price P_t^w. Define $\mu_t = P_t/P_t^w$ as the retail-price markup. To hire labor, wholesale firms post vacancies V_t, interview and screen applicants, and make retention decisions. Expressed in terms of final retail goods, the current surplus of a firm-worker match involving a high-efficiency worker is

$$s_t^h = \left(\frac{\phi^h h_t^h}{\mu_t}\right) - \frac{v(h_t^h)}{\lambda_t} - w_t^{u,h} + q_t^h, \tag{5}$$

where h_t^h denote hours worked by an employed high-efficiency worker (so a high-efficiency worker produces $\phi^h h_t^h$ of the wholesale good), $v(h_t^h)$ is the disutility of hours worked, λ_t is the marginal utility of consumption, $w_t^{u,h}$ is the value of an unmatched high-efficiency worker's outside opportunity, and q_t^h is the expected continuation value of a match with a high-efficiency worker. All type h workers will work the same hours since they have the same productivity, and h_t^h is chosen optimally to maximize the match surplus. This implies

$$v_h'(h_t^h) = \phi^h \left(\frac{\lambda_t}{\mu_t}\right) \tag{6}$$

The surplus of a match involving a low-efficiency worker is

$$s_{i,t}^l = \left(\frac{a_{i,t}^l \phi^l h_{i,t}^l}{\mu_t}\right) - \frac{v(h_{i,t}^l)}{\lambda_t} - w_t^{u,l} + q_t^l. \tag{7}$$

Hours will again be chosen to maximize the surplus and will thus vary with a type l worker's idiosyncratic productivity realization to satisfy

$$v_h'(h_{i,t}^l) = a_{i,t}^l \phi^l \left(\frac{\lambda_t}{\mu_t}\right). \tag{8}$$

If a low-efficiency worker's productivity is too low, the surplus will be negative, leading to endogenous separation (or screening in the case of an interviewed job seeker). From (7), the cutoff value of worker productivity at which the surplus produced by a low-efficiency worker equals zero is

$$\bar{a}_t = \frac{\mu_t \left(w_t^{u,l} + \frac{v(\hat{h}_t^l)}{\lambda_t} - q_t^l\right)}{\phi^l \hat{h}_t^l}, \tag{9}$$

where $\hat{h}_t^l$ maximizes the joint surplus for a worker with $a_{i,t}^l = \bar{a}_t$. Equation (9) implies that $\bar{a}_t$ is the same for all firm considering the retention or hire of a low–efficiency worker. If the retail price markup μ_t increases, $\bar{a}_t$ will rise, lowering the fraction of low-efficiency job seekers who receive job offers and increasing the endogenous separation rate of already employed low efficiency workers.[7] Low efficiency workers then become a larger fraction of the unemployed pool. Also, because the low efficiency workers lose jobs faster and have a harder time finding new employment since they are more likely to be screened out during the interview process, the average duration of unemployment increases.

Our assumption that the idiosyncratic productivity shocks are serially uncorrelated implies

[7] If we had included an aggregate producitivity shock z_t, then the denominator of (9) would become $z_t \phi^l \hat{h}_t^l$ and an increase in z_t would decrease $\bar{a}_t$.

the continuation value q_t^j depends on the efficiency-type of the worker in a match but is the same for all matches of the same efficiency-type. These continuation values are given by

$$q_t^h = \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[(1-\rho^x) s_{t+1}^h + w_{t+1}^{u,h} \right]. \tag{10}$$

and

$$q_t^l = \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[(1-\rho^x) \int_{\bar{a}_{t+1}}^{1} s_{i,t+1}^l f(a_i^l) da_i^l + w_{t+1}^{u,l} \right]. \tag{11}$$

We assume wages are determined by Nash bargaining with the worker receiving a constant share η of the match surplus. The value of unemployment is w^u, the marginal productivity of an unemployed worker in home production (see below), plus the expected probability of being employed and receiving the surplus share ηs_{t+1}^j plus the expected value of remaining unemployed. For a high-efficiency worker,

$$w_t^{u,h} = w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left(k_{t+1}^w \eta s_{t+1}^h + w_{t+1}^{h,u} \right), \tag{12}$$

while for a low-efficiency worker,

$$w_t^{u,l} = w^u + \beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \right) \left[k_{t+1}^w \eta \int_{\bar{a}_t+1}^{1} s_{i,t+1}^l f(a_i^l) da_i^l + w_{t+1}^{u,l} \right]. \tag{13}$$

With firms receiving a share $1-\eta$ of the surplus from a match, the job posting condition is

$$k_t^f (1-\eta) \left[(1-\gamma_t) s_t^h + \gamma_t (1-\rho_t^n) E_t(s_{i,t}^l | hiring) \right] = \kappa, \tag{14}$$

where κ is the cost of posting a vacancy, expressed in terms of final goods. The left side of (14) is the probability of an interview k_t^f times the firm's share of the expected surplus, since with probability $(1-\gamma_t)$ the firm interviews (and hires) a high-efficiency worker and with probability γ_t, it interviews a low-efficiency worker. Because the expected surplus from a high efficiency worker is greater than the expected surplus obtained from entering into an interview with a low efficiency worker, the incentive to post vacancies falls when a rise in γ_t reduces the average quality of the unemployment pool.

Output of wholesale goods is obtained by aggregating over the output produced by employed high-efficiency workers and the output produced by employed low-efficiency workers (i.e., those with idiosyncratic productivity levels greater than $\bar{a}_t$):

$$Q_t = \left\{ (1-\xi_t)\phi^h h_t^h + \xi_t \phi^l \left[\frac{\int_{\bar{a}_t}^1 a_{i,t}^l h_{i,t}^l dF(a_i^l)}{1 - F(\bar{a}_t)} \right] \right\} N_t. \tag{15}$$

1.2 Households, retail firms, monetary policy, and market clearing

The rest of the model can be briefly summarized as it follows the standard specification in new Keynesian models.

Households: The representative household maximizes

$$\mathrm{E}_t \sum_{i=0}^{\infty} \beta^i \delta_t^i \left\{ D_t \frac{\mathcal{C}_{t+i}^{1-\sigma}}{1-\sigma} - \left[v(h_{t+i}^h)(1-\xi_{t+i})N_{t+i} + \xi_{t+i} N_{t+i} \int_{\bar{a}_t}^1 v(h_{i,t+i}^l) f(a) da \right] \right\}, \tag{16}$$

where $\sigma > 0$ is the coefficient of relative risk aversion, δ_t^i is a discount rate shock, D_t is an aggregate preference shock, $\mathcal{C}_t$ is the sum of a market-purchased composite consumption good C_t and home-produced consumption by unemployed workers $C_t^u = (1 - N_t)w^u$. In (16), the term

$$v(h_{t+i}^h)(1-\xi_{t+i})N_{t+i} + \xi_{t+i} N_{t+i} \int_{\bar{a}_t}^1 v(h_{i,t+i}^l) f(a^l) da^l$$

is the disutility to the household of having N_t members working, where hours worked depends on type and the idiosyncratic productivity shocks. We assume $v(h_{t+i}) = \ell h_{t+i}^{1+\chi}/(1+\chi)$.

Market consumption C_t is a Dixit-Stiglitz composite good consisting of the differentiated products produced by retail firms and is defined as

$$C_t = \left[\int_0^1 c_{k,t}^{\frac{\theta-1}{\theta}} dk \right]^{\frac{\theta}{\theta-1}} \qquad \theta > 0.$$

Given prices $p_{k,t}$ for the final goods, this preference specification implies the household's demand for good k is

$$c_{k,t} = \left(\frac{p_{k,t}}{P_t} \right)^{-\theta} C_t, \tag{17}$$

where the aggregate retail price index P_t is defined as

$$P_t = \left[\int_0^1 p_{k,t}^{1-\theta} dj \right]^{\frac{1}{1-\theta}} .$$

If i_t is the nominal rate of interest, the representative household's first order conditions imply the following must hold in equilibrium:

$$\lambda_t = \beta(1 + i_t)\mathrm{E}_t \left(\frac{P_t}{P_{t+1}} \right) \lambda_{t+1}. \tag{18}$$

Retail firms: There are a continuum of retail firms, indexed by j, who purchase the wholesale good and convert it into differentiated final goods that are sold to households and wholesale firms. Retail firms maximize profits subject to a CRS technology for converting wholesale goods into final goods, the demand functions (17), and a restriction on the frequency with which they can adjust their price. Each period a firm can adjust its price with probability $1 - \omega$. The real marginal cost for retail firms is the price of the wholesale good relative to the price of final output. This is just the inverse of the markup of retail over wholesale goods:

$$\mu_t \equiv \frac{P_t}{P_t^w}$$

This retail price markup is the driving force for inflation.

Monetary policy: Monetary policy is represented through a simple Taylor-type instrument rule, constrained by the requirement that the nominal rate be non-negative. The specific rule we incorporate takes the form

$$\ln(1 + i_t) = \max \left[0, \ - \ln \beta + \chi_i \ln(1 + i_{t-1}) + (1 - \chi_i) \left[\omega_\pi \pi_t + \omega_y \left(\ln Y_t - \ln \overline{Y} \right) \right] + \varepsilon_t \right] . \tag{19}$$

where ε_t is an i.i.d. policy shock. As a baseline policy we assume an inflation-targeting policy, setting $\omega_\pi = 1.5$, $\omega_y = 0$, $\chi_i = 0$.

Market clearing: Goods market clearing requires that household consumption of market produced goods equals the output of the retail sector minus final goods purchased by wholesale firms to cover the costs of posting job vacancies, or

$$Y_t = \Delta_t \left(C_t + \kappa V_t \right), \tag{20}$$

where $\Delta_t \geq 0$ is a measure of relative price dispersion.

2 Inefficient separations

Before carrying out our quantitative exercises to assess the channels introduced by heterogeneity-driven selectivity in hiring and separations, we address the implications of worker heterogeneity for the efficiency of the competitive equilibrium.

In a basic new Keynesian model with search and matching frictions but a homogeneous labor force, Ravenna and Walsh (2011) identify four distortions that arise in the competitive equilibrium with Nash bargaining over wages. The first two are standard in new Keynesian models: a non-zero steady-state markup under monopolistic competition generates a level of output that is inefficiently low, and price (or wage) rigidity generates inefficient fluctuations in the markup. The third distortion arises because the markup affects the optimal choice of hours and so fluctuations in the markup distort hours from their efficient level. Finally, the fourth distortion occurs when the Hosios condition is not satisfied, causing the vacancy posting condition in the competitive equilibrium to be distorted relative to the social planner's allocation. These four distortions would be eliminated if (a) a subsidy to firms is used to raise steady-state output to its efficient level; (b) the markup is constant (i.e., prices are stable); and (c) the Hosios condition holds.

The existence of time-varying worker heterogeneity implies an additional distortion. When firms separate or screen out low-efficiency workers at the interview stage, they also jointly determine the average efficiency level of the pool of searching workers from which all new matches are formed. This last effect represents an externality that is not internalized by a profit-maximizing firm. The distortion that results is not eliminated even when the Hosios condition is met.

To compare the social planner's allocation with that of the competitive market, let $\bar{s}^j_{i,t}$

be the surplus value of an employed worker i of type j in the planner's allocation. Because high-efficiency workers are identical, the i notation is unnecessary when $j = h$. Recall that $s^j_{i,t}$ denotes the joint worker-firm surplus for a match involving worker i of type j in the competitive equilibrium.

Suppose the steady-state markup distortion has been eliminated, price stability holds, and the Hosios condition holds. Then $\mu_t = 1$ and $\eta = \alpha$. In this case, the surplus for a match with a type h worker is[8]

$$s^h_t = \left[\left(\frac{\chi}{1+\chi}\right)\left(\phi^h\right)^{\frac{1+\chi}{\chi}}\left(\frac{\lambda_t}{\ell}\right) - w^u\right] + (1-\rho^x)\beta \mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k^w_{t+1}\right)s^h_{t+1}.$$

Let the surplus for a type h match in the efficient allocation be denoted by $\bar{s}^h_t$. The appendix shows that

$$\begin{aligned}\bar{s}^h_t &= \left[\left(\frac{\chi}{1+\chi}\right)\left(\phi^h\right)^{\frac{1+\chi}{\chi}}\left(\frac{\lambda_t}{\ell}\right) - w^u\right] + (1-\rho^x)\,\beta \mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k^w_{t+1}\right)\bar{s}^h_{t+1} \\ &\quad -(1-\alpha)\,\beta \mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\gamma_{t+1}k^w_{t+1}X_{t+1}, \qquad (21)\end{aligned}$$

where the new variable X_{t+1} in the planner's allocation is defined as

$$X_{t+1} \equiv (1-\rho^x)\left\{\bar{s}^h_{t+1} - \left[1-F(\bar{a}_{t+1})\right]\bar{s}^l_{t+1}\right\} \geq 0.$$

This expression is equal to the average surplus at $t+1$ of a high-efficiency worker, conditional on surviving the exogenous separation hazard, minus the average surplus at $t+1$ of a low-efficiency worker, conditional on that worker surviving the exogenous separation hazard and being retained. From the perspective of the social planner, the value of having an high-efficiency worker employed at time t consists of the current value of production net of w^u, the first term on the right in (21), plus the probability of still having a high-efficiency worker employed the following period, the second term in (21). This second terms accounts for the

[8] This uses the fact that the optimal hours for worker i are $v'/\lambda_t = \ell\left(h^l_{i,t}\right)^\chi = a_{i,t}\phi^j/\mu_t$ and therefore

$$\left(\frac{a_{i,t}\phi^j h^j_{i,t}}{\mu_t}\right) - \frac{v(h^l_{i,t})}{\lambda_t} = \frac{\chi}{1+\chi}\left(\frac{\lambda_t}{\ell}\right)^{\frac{1}{\chi}}\left(\frac{a_{i,t}\phi^j}{\mu_t}\right)^{\frac{1+\chi}{\chi}}.$$

fact that some type-h workers who exogenously separate may be rehired and produce in period $t+1$.[9] The third term arises because the employment of an additional high-efficiency worker lowers the average productivity of the pool of job seekers, making it more likely that a new hire would be a low-efficiency worker. This "cost" equals the probability such a hire survives the exogenous separation hazard times the surplus of a type h minus that of a type l, adjusted for the endogenous separation probability for a type l.

Evaluated at the social planner's allocation, the difference between the private and efficient level of surplus is given by

$$s_t^h - \bar{s}_t^h = (1-\rho^x)\,\beta \mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k_{t+1}^w\right)\left(s_{t+1}^h - \bar{s}_{t+1}^h\right) + \mathrm{E}_t \gamma_{t+1} Z_{t+1} \tag{22}$$

where

$$Z_{t+1} = (1-\alpha)\,\beta\left(\frac{\lambda_{t+1}}{\lambda_t}\right) k_{t+1}^w X_{t+1} \geq 0.$$

Recursively solving (22) forward implies that $s_t^h - \bar{s}_t^h$ can be expressed as the present discounted value of expected current and future expected values of $\mathrm{E}_t \gamma_{t+1} Z_{t+i}$, which is nonnegative. Thus, $s_t^h - \bar{s}_t^h \geq 0$, and matches involving type h workers generate an inefficiently *high* surplus in the competitive equilibrium relative to the first-best allocation. Only if workers are homogenous and all of high-efficiency so that $\gamma_{t+i} = 0$ would $s_t^h = \bar{s}_t^h$.

For type l workers, the opposite condition will hold. For an employed type l worker i,

$$s_{i,t}^l - \bar{s}_{i,t}^l = (1-\rho^x)\,\beta \mathrm{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\right)\left(1-\alpha k_{t+1}^w\right)\left[1 - F(\bar{a}_{t+1})\right]\left(s_{t+1}^l - \bar{s}_{t+1}^l\right) - \mathrm{E}_t\left(1-\gamma_{t+1}\right) Z_{t+1}. \tag{23}$$

Thus, $s_{i,t}^l < \bar{s}_{i,t}^l$. Type l workers generate a *smaller* surplus in the competitive equilibrium than in the efficient allocation unless the worker force consists entirely of low-efficiency workers so that $\gamma_t = 1$.

These differences in valuations translate into differences in the optimal cutoff productivity level $\bar{a}_t$. Specifically, both $s_{i,t}^l$ and $\bar{s}_{i,t}^l$ are increasing functions of the idiosyncratic productivity of worker i. Since private matches involving type l workers end whenever $a_{i,t}^l < \bar{a}_t$, with $\bar{a}_t$

[9] This accounts for the appearance of the αk^w term.

defined such that $s^l_{i,t}(\bar{a}_t) = 0$, the fact that $\bar{s}^l_{i,t}(a_{i,t}) \geq s^l_{i,t}(a_{i,t})$ implies

$$\bar{s}^l_{i,t}(\bar{a}_t) \geq s^l_{i,t}(\bar{a}_t) = 0.$$

That is, a worker who generates a zero surplus in the competitive equilibrium still generates a positive surplus in the social planner's allocation. Therefore, $\bar{a}^{sp}_t \leq \bar{a}_t$, where $\bar{a}^{sp}$ is the cutoff productivity level in the efficient allocation (i.e., such that $\bar{s}^l_t(\bar{a}^{sp}_t) = 0$). Thus, some type l workers who experience endogenous separation and become unemployed in the competitive equilibrium would remain employed in the efficient allocation. Similarly, some unemployed type l workers who obtain interviews but are screened out in the competitive equilibrium would be hired in the efficient equilibrium. This translates into a higher share of low-efficiency workers among unemployed workers, and a lower expected benefit to posting vacancies. Ceteris paribus, endogenous separations are too high in the competitive equilibrium, average unemployment is also too high, and average unemployment duration is inefficiently long.

The appendix shows that this result can be extended to the case in which both worker types experience individual-specific i.i.d. productivity shocks that can only be observed by the firm if the worker is interviewed for a job or is a current employee. To focus on average productivity differences while holding the dispersion of idiosyncratic productivity realizations within types the same, assume the productivity of worker i of type j is $a^j_{i,t} = \phi^j + e_{i,t}$ for $j = h, l$, where ϕ^j is the expected productivity of a worker of type j. By definition, $\phi^h > \phi^l$; on average, type h workers are more productivity than type l workers. In turn, $e_{i,t}$ is an idiosyncratic component of worker productivity drawn from the same distribution F for both types, with mean zero and variance σ^2_e. The cutoff productivity realization that triggers endogenous separation (or screening out in the interview process) for type j is $\bar{a}^j_t$. If the endogenous separation rates is denoted $\rho^{n,j}_t$ for type $j = l, h$,

$$\rho^{n,h}_t \leq \rho^{n,l}_t;$$

The endogenous separation rate for type h workers is lower than for type l workers (see appendix).

For the social planner's problem, (22) is unchanged, but X_{t+1} is now equal to

$$X_{t+1} \equiv (1-\rho^x)\left[\left(1-\rho^{n,h}_{t+1}\right)\bar{s}^h_{t+1} - \left(1-\rho^{n,j}_{t+1}\right)\bar{s}^l_{t+1}\right] \geq 0.$$

It then holds that $s^h_{i,t} - \bar{s}^h_{i,t} \geq 0$ and $s^l_{i,t} - \bar{s}^l_{i,t} \leq 0$ as before. Therefore, $\bar{a}^{sp,h}_t \geq \bar{a}^h_t$ and $\bar{a}^{sp,l}_t \leq \bar{a}^l_t$ Firms retain and hire too many high-efficiency workers and separate from and screen out too many low-efficiency workers.

Because firms ignore the effect of their hiring and retention decisions on the composition of the pool of job seekers, they reject (fire) too many low productivity workers and hire (retain) too many high productivity workers, even when the standard distortions in a search and matching labor market with Nash bargaining and sticky prices are eliminated. Thus, the inflow into unemployment among type l workers is inefficiently high and the outflow into employment is inefficiently low. This results in an unemployment rate for type l workers that is inefficiently high. The reverse holds for type h workers, and the unemployment rate for these workers is inefficiently low.

3 Aggregate demand shocks

In this section, we discuss the baseline model parameterization and then illustrate the implications of selection for the economy's response to an aggregate demand shock. Both demand and layoff shocks, considered in section 4, will play roles in our pandemic scenario in section 5.

3.1 Parameterization

The value of home production w^u, the coefficient ℓ scaling the disutility of labor hours, the cost of vacancy posting κ, the productivity of the matching technology ψ, the relative steady state productivity of high to low-efficiency workers $\phi^h / \left(\phi^l \int_0^1 a_i dF(a_i)\right)$ and the labor force share of low-efficiency workers $\bar{\gamma}$ are jointly chosen to match the steady-state values for six variables with average aggregate data. Table 1 reports the matched steady state values, together with the additional parameters used in the numerical simulations.

The steady state unemployment rate is obtained averaging BLS quarterly data over 1948Q to 2019Q4. Since we do not have a direct measure for the unemployment rate of workers sorted according to unobservable productivity differentials, we measure the unemployment rate of workers with different efficiency levels using age-related data. We identify unemployment rates for low and high-efficiency workers with rates for workers' age-groups 16 to 24 and over-24. The steady state hours per worker h_{ss}^{av}, the steady state aggregate separation rate and the probability of a match between an applicant and a vacancy k_{ss}^{f} are parameterized to standard values in US business cycle literature. The share of output devoted to hiring activities is in line with empirical evidence reported in Ravenna and Walsh (2008). Jointly, these targeted moments return a matching function productivity ψ of 0.70 and a vacancy posting cost κ of 0.08.

Table 2 reports the implied steady-state values for labor market variables and productivity. The parameterization implies that $\bar{\gamma}$, the steady-state share of l workers in the labor force L, is 23.3%. Because the separation rate of l workers is about 45% larger than the overall separation rate, their share γ_{ss} in the steady-state pool of job seekers is 30%, while their share ξ_{ss} in the steady-state employment pool is 22%. Thus, low-efficiency workers are over-represented in the pool of unemployed, and this pool has a lower average productivity than the pool of employed workers. When entering into a match with a firm, high-efficiency workers are expected to have an hourly productivity 18% higher than low-efficiency workers. The extent of selection at hiring is small: firms expect to turn down at an interview only about 1% of the workers they randomly match with (or 3.1% of the interviewed l workers). The screening-out rate increases in a recession for two reasons: the unemployment share of low-efficiency workers increases, and their minimum productivity (i.e., $\bar{a}_t$) has to raise considerably for a match to generate a positive value.

3.2 Demand shocks and time-varying heterogeneity

We capture the impact of a fall in aggregate demand by simulating the economy's response to a negative realization of the preference shock D_t. The shock is calibrated to produce a 1% fall in output. The basic responses are as expected in the face of a demand shock; output, inflation, employment, hours, and labor market tightness all fall, and unemployment rises. Our focus is on how this shock affects the two worker types differentially. The decline in demand induces firms to reduce employment, but normal exogenous separations are insufficient to achieve this reduction. Thus, endogenous separations rise as firms become more selective ($\bar{a}_t$ rises) and more matches with low-efficiency workers end. The inflow into unemployment is therefore a selective process that results in a shift in the average productivity of the unemployed. The increased selectivity of firms will also be at work in the recovery, as job finding rates for low-efficiency workers fall.

The dashed lines in figure 2 show the effects of the negative demand shock on the aggregate unemployment rate (top panel), together with the effects on the unemployment rates for the low- and high-efficiency workers (middle and bottom panels). The distributional impact of the recession across worker types is shown clearly in the large differences in unemployment experienced by the two worker types. The impact of labor market slack falls primarily on the low-efficiency workers even though they are a small share of steady-state employment. The in the upper left panel (dashed line) of figure 3 shows the fall in labor market tightness, while the upper right panel (dashed line) shows that the share of low-efficiency workers among the pool of job seekers, γ_t, remains persistently above its steady-state value for a number of quarters.

The persistence in unemployment among the low-efficiency workers results from their lower job finding rate. The lower left panel of figure 3 show the job finding rate for both worker types, as well as the aggregate average job finding rate. For high-efficiency workers, the job finding rate is equal to the interview rate k_t^w. This falls as the number of vacancies per job seeker falls. For low-efficiency workers, the job finding rate equals $(1 - \rho_t^n)\, k_t^w$, as a low-efficiency workers needs to have a current productivity realization greater than $\bar{a}_t$ to be hired, and this occurs with probability $1 - \rho_t^n$. The job-finding rate for low-efficiency workers declines due to the fall in the likelihood of being interviewed as labor market tightness falls, an effect shared with high-efficiency workers, but it also declines because the fall in demand makes firms more selective; the cutoff productivity level $\bar{a}_t$ rises, reducing $1 - \rho_t^n$. The lower right panel of the figure shows the job filling rate rises, as would be expected from the fall in labor market tightness.

The fall in demand causes the price of wholesale goods (the flexible-price sector) to fall, but with sticky retail prices, the retail price markup increases, lowering marginal costs for retail firms and reducing inflation. The rise in μ_t lowers the value of wholesale output in terms of final goods, increasing $\bar{a}_t$ (see 9). Hence, endogenous separations rise, increasing the fraction of low-efficiency workers among job searchers.

The impact of the changing composition of the unemployment pool on optimal vacancy posting can be illustrated by rewriting the vacancy posting condition (14) as

$$\theta_t = \frac{V_t}{S_t} = \left\{ \frac{\psi}{\kappa}(1-\eta) \left[(1-\gamma_t) s_t^h + \gamma_t (1-\rho_t^n) E_t(s_{i,t}^l | hiring) \right] \right\}^{1/\alpha}.$$

Labor market tightness, θ_t, depends on the expected surplus from a match. Since the surplus from a high-efficiency worker is higher than the expected surplus obtained from entering into an interview with a low efficiency worker, the fall in average productivity among job seekers as γ_t increases reduces the efficiency-weighted expected surplus and thereby reduces the incentive to post vacancies. Thus, the larger the increase in the share of l workers among job seekers, the larger is the fall in the number of vacancies per searching worker, and the larger is the fall in the interview rate $k_t^w = \psi \theta_t^{1-a}$.

Sizeable movements in the average labor productivity of the unemployed, driving the vacancy-posting incentives of firms, are obtained even though the baseline calibration actually incorporates very little heterogeneity across workers. Our parameterization assumes low-efficiency workers are only 23.3% of the labor force, and the average TFP of the employed worker-hour is only 1.95% higher than the average TFP for the unemployed. Changes in the relative quality of job seekers and employed workers arises because separations and hires are not random. Separations occur selectively, as not all workers face the same separation rate, and hires are done selectively, as not all workers face the same probability of being hired. Selectivity contrasts with a search and matching model with only exogenous separations and random matching in which all workers face the same separation probability and in which all matches translate into hires.

These endogenous movements in productivity cause fluctuations in employment and output to be decoupled in ways that would not occur if all workers were identical. Effectively, in this model, employed workers are not representative of unemployed workers. Since the compositional changes in the flow into unemployment during a recession are large relative to the compositional shares in the stock of unemployed, they can affect aggregate labor market variables. In a model with homogeneous workers and endogenous separation, this amplification mechanism would not work, since average worker productivity would move identically for job seekers and employed workers.

Importantly, section 2 showed that even when the Hosios condition and price stability both hold, the rise in $\bar{a}_t$ that leads to persistent and long-lasting spells of unemployment for low-efficiency workers is inefficiently large. The labor market slack resulting from the negative demand shock falls too heavily on low-efficiency workers, while unemployment among high-efficiency workers rises too little. And the flow of low-efficiency workers into the unemployment pool reduces the incentive for firms to post vacancies and thereby also lowers the job-finding probability of high-productivity workers.

4 Shocks to exogenous separations

A key characteristic of the increase in unemployment generated by the COVID-19 pandemic in early 2020 were the mass layoffs that affected workers of all types. By causing a large inflow of both worker types into unemployment, a mass layoff may have a much smaller impact on the composition of the pool of unemployed than the endogenous separations caused by the demand shock analyzed in section 3. In this section, therefore, we consider a shock to exogenous separations that affects both worker types. Then, in the following section, we investigate a COVID-19 scenario in which the economy is subject to both a negative demand shock and a positive shock to exogenous separations.

The size of the ρ^x shock is calibrated to produce the same 1% decline in output that was generated by the negative demand shock. Because the exogenous separation hazard affects all workers, regardless of type, we view this experiment as illustrating the effects of a non-selective increase in the inflow into unemployment. This contrasts with the demand shock in which initial separations rose because of the selective ending of matches with low-efficiency workers. Importantly, we find that the apparent non-selective initial rise in unemployment leads to endogenous adjustments that cause firms to become more selective. This results in low-efficiency workers experiencing both increased exogenous separations and increased inflow into unemployment from endogenous separations. They also suffer longer spells of unemployment. Overall, the impact of a burst in random separations ends up being heavily biased towards raising joblessness for low-efficiency workers.

The solid lines in figure 2 show the paths of the average unemployment rate in response to the exogenous separation shock and the rates for low- and high-efficiency workers. Even though the demand and separation shocks are calibrated to produce the same fall in output, the employment consequences of the increase in exogenous separations are much worse for both worker types than with the demand-driven recession. The increase in the exogenous separation shock has two effects. First, it increases the flow from employment into unemployment. However, firms would have an incentive to re-hire the excess unemployed, since neither a demand shock nor a productivity shock have changed households demand for goods. Second, it lowers the value of any match, which now have a lower expected duration. To obtain a 1% drop in output, the exogenous increase in the separation rate has to be sufficiently large, resulting in a large fall in employment, while hours per worker increase.

Importantly, for low-efficiency workers the positive shock to exogenous separations also leads to a rise in endogenous separations. This causes the total rise in unemployment to be concentrated among the low-efficiency workers, as seen by the large rise in γ_t shown in the upper right panel of figure 3. While the initial shock to ρ^x affects both worker types, it still affects the aggregate composition of the pool of job seekers. The induced rise in ρ_t^n implies firms have become more selective, and this amplifies the decline in job finding rates for low-efficiency workers - similarly to the case of a demand shock, shown in the lower left panel of figure 3 - so that low-effficiency workers experience a rise in the inflow into unemployment and a fall in the outflow.

High-efficiency workers are also affected. The rise in the number of job seekers reduces the job-finding probability of high-productivity workers, and the adverse impact of a rise in γ_t on the average productivity of job seekers dampens the incentive for firms to post vacancies. However, it is low-efficiency workers that are most affected. While k^w, the job finding rate for high-efficiency workers falls by about 3.5 percentage points, compared to a 1 percentage point fall with the demand shock, $k^{w,l}$ falls by over 8 percentage points, more than three times the decline observed following a demand shock. Furthermore, the lower right panel of the figure shows that, after an initial rise in the job filling rate due to the decline in labor market tightness shown in the upper left panel, firms find it harder to fill jobs. The increasing share of the unemployed who are low-efficiency begins one period after the shock, implying that more interviews are with low-efficiency workers, and the rise in the cutoff productivity level $\bar{a}_t$ means fewer of these interviews result in successful matches as more workers are screened out during interviews.[10]

Why do firms become more selective after a rise in ρ^x? Cutoff productivity is defined such that the joint surplus for a low-efficiency worker is zero. Using the first-order condition for hours, the surplus from a match with a type l worker with productivity $a^l_{i,t}$ is[11]

$$s^l_{i,t} = \left[\left(\frac{\chi}{1+\chi}\right) \left(\frac{a^l_{i,t}\phi^l}{\mu_t}\right)^{\frac{1+\chi}{\chi}} \left(\frac{\lambda_t}{\ell}\right) - w^u \right] + (1-\rho^x)\beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left(1-\eta k^w_{t+1}\right) \left(1-\rho^n_{t+1}\right) s^h_{t+1}.$$

The cutoff value $\bar{a}_t$ at which $s^l_{i,t} = 0$ is given by the solution to

$$\left(\frac{\bar{a}_t}{\mu_t}\right)^{\frac{1+\chi}{\chi}} = \frac{1+\chi}{\chi} \left(\frac{\ell}{\lambda_t}\right)^{\frac{1}{\chi}} \left[w^u - (1-\rho^x)\beta \mathrm{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t}\right) \left(1-\alpha k^w_{t+1}\right) \left(1-\rho^n_{t+1}\right) s^l_{t+1} \right]. \quad (24)$$

$\bar{a}_t$ depends on μ_t, λ_t, and the continuation value of a surviving match. The markup initially rises as wholesale prices fall relative to the retail price level. This lowers the real marginal value of match output, so a worker must be more productive to generate a positive surplus. This channel increases $\bar{a}_t$. The continuation value of a job is given by the second term within the brackets on the right in (24). The rise in ρ^x reduces this value and acts to raise selectivity (increase $\bar{a}_t$) as a worker's current productivity must be higher to generate a positive surplus when the continuation value falls. Offsetting these effects is the fall in the expected interview rate k^w_{t+1}, which acts to lower the continuation value, and the rise in the marginal utility of consumption λ_t as total consumption falls. These effects act to decrease $\bar{a}_t$. For our baseline calibration, the net effect of the exogenous separation rate shock is to raise $\bar{a}_t$. The rise in ρ^n_t helps to explain the much larger increase in the share of low-efficiency workers in the pool of

[10] In the period of the mass-layoff shock, γ_t actually falls as the inflow is dominated by the type h workers because they comprise a larger share of employment.

[11] This uses the fact that the optimal hours for worker i are $v'/\lambda_t = \ell \left(h^l_{i,t}\right)^\chi = a^l_{i,t}\phi^l/\mu_t$ and therefore

$$\left(\frac{a^l_{i,t}\phi^l h^l_{i,t}}{\mu_t}\right) - \frac{v(h^l_{i,t})}{\lambda_t} = \frac{\chi}{1+\chi} \left(\frac{\lambda_t}{\ell}\right)^{\frac{1}{\chi}} \left(\frac{a^l_{i,t}\phi^l}{\mu_t}\right)^{\frac{1+\chi}{\chi}}.$$

job seekers that occurs in the face of the exogenous separation shock.

The adjustment of hours after a positive ρ^x shock is quite different than after a negative demand shock. Hours fall after the demand shock, as total hours decline along both the extensive and the intensive margins. In contrast, the exogenous separation shock destroys matches along the extensive margin, but for the surviving matches, hours rise for both high-efficiency and low-efficiency workers. For employed high-efficiency workers, the fall in output increases the marginal utility of consumption, making it optimal for the worker to substitute towards supplying more hours as the value of leisure relative to consumption declines.

For low-efficiency workers, this same substitution effect is at work. But a second effect is also in play. Optimal hours worked by low-efficiency worker i equal

$$h^l_{i,t} = \left(\frac{a^l_{i,t} \phi^l \lambda_t}{\ell \mu_t} \right)^{\frac{1}{\chi}},$$

which is increasing in the worker's idiosyncratic productivity $a^l_{i,t}$. Because $\bar{a}_t$ has increased, low-efficiency workers who remain employed are now, on average more efficient. This shift in the distribution of productivity among low-efficiency workers results in an increase in the average hours of employed low-efficiency workers.

To summarize, a shock to exogenous separations – a mass layoff – initially affects all employed workers in a non-selective fashion. But the endogenous adjustments that result from this shock lead to further increases in unemployment for low-efficiency workers as selectivity generates more endogenous separations and greater screening out of those low-efficiency job seekers who do get interviewed. Such workers see a fall in their job finding rate and an increase in the length of unemployment spells. As noted previously, even when standard distortions in new Keynesian models with search and matching frictions are eliminated, the low-efficiency workers bear an inefficiently large share of the unemployment.

While the separations shock was calibrated to produce the same fall in output as the demand shock discussed in section 3, the effects on unemployment are much larger as hours move quite differently in the face of the two shocks. They fall after the demand shock and increase after the separation shock. Given that the output decline is the same in both cases, an increase in hours after the separation shock implies a larger fall along the extensive margin then seen with the demand shock. These differences in employment levels lead to quite different movement in average output per worker. The demand shock causes output to fall by 1% and employment to fall by at most 0.5%; output per worker therefore falls by 0.5%. In sharp contrast, a separation shock reduces output by 1% while employment declines by 2.5%; output per worker *rises* by 1.5%.

5 A COVID-19 recession and the impact of selection

Beginning in March 2020, the COVID-19 pandemic caused both a collapse in demand *and* a surge in separations in the U.S. The resulting rise in unemployment has been both large and common across industries and states. According to BLS data, by the second half of May 2020 around 39 million workers had lost their jobs across the United States, with the official unemployment rate jumping from 4.4% in March to 14.7% in April before receding slightly to 13.3% in May and falling further to 11.1% in June, though the number of permanent job losers continued to increase in June. Survey estimates in May were predicting the unemployment rate would climb to 20% by June 2020, with a fall in working hours per employee of 12.5%, and a decline in earnings per worker of 21%, mostly as a consequence of fewer hours worked (Bick and Blandin 2020; Barrero, Bloom, and Davis 2020). Across U.S. states, the unemployment rates increases related to COVID-19 have ranged from nearly 40% to slightly less than 8%. Similarly, job loss has occurred across all industrial sectors of the economy; only mining and agriculture saw year-over-year increases in unemployment from May 2019 to May 2020 of less than 300%.

To describe the COVID-recession, we assume that the economy is hit by both a negative demand shock and a positive separation rate shock. First, shelter-in-place and social distancing requirements result in excess separations that we capture through a shock that increases exogenous separations. Second, we capture the accompanying fall in aggregate demand through a negative preference shock. Thus, our benchmark COVID scenario combines the two shocks that we discussed in sections 3 and 4.

Our approach acknowledges that the dynamics of the model economy are driven by deep shocks that may be unobservable. Thus, we build hypothetical scenarios based on observable variables, and let the model back out the shocks that drive the dynamics, producing conditional forecast for labor market variables and aggregate variables which are not supplied to the model. Specifically, we ask the model to match a given forecast path for total separations and output. The model then endogenously allocates the total separations across an exogenous, unexpected shock to ρ^x and an endogenous portion driven by selection. Similarly, the model will back out the exogenous shift in household preferences consistent with the paths for output and total separations.

To produce the COVID scenario, we assume an output path in the conditional forecast exercise calibrated to reach a trough that is 10% below trend in the third quarter before rebounding to 3% below trend in the sixth quarter. For the path for total separations, we assume it increases rapidly and peaks in the third quarter at 100% of its steady state value. It then falls to half that peak increase by the fourth quarter before subsequently falling back to steady state within the following six quarters.[12]

Both the demand and separations shocks are essential to discuss a pandemic-induced recession. The exogenous separation shocks acts as a supply shock; the value of each existing or potential match falls, ceteris paribus, since the cost per vacancy is fixed but its return in terms of match-lifetime production is lower, given that the match is expected to have a shorter duration. The preference shock acts as a demand shock that can match the output loss for a given separation path. By itself, the separation shock is not sufficient to lower demand in line with current expectations for the U.S. economy; the shock to exogenous separations, if not coupled with a fall either in demand or TFP, pushes firm to rehire a large share of the separating workers.

Our baseline model captures the recession and recovery path in an economy with permanent separations and where labor separations require that firm-workers matches be re-established through a costly matching process.[13] We show that this results in an L-shaped recovery in the labor market. To assess the role of selection in this adjustment process, we contrast our model with the implications under an alternative parameterization in which labor flows are very large

[12] The JOLTS survey documents that the private sector separation rate for the private sector increased from a monthly rate of 4% (the value reached in February 2020, approximately equal to the value for March 2019) to 11.1% in March 2020 and 8.8% in April 2020, with more than half of the separations in each month from March to June being classified as temporary. We consider the impact of temporary layoffs in section 7.

[13] We take up the issue of temporary layoffs in section 7.

every period.[14] While selection is present in gross labor flows, its cyclicality is negligible. Such an economy mimics the outcomes that would occur in the absence of any selection. Under this parameterization – which we label no-selection for short – the recessionary increase in the total separation rate is generated by a much larger increase in the exogenous separation shock, since endogenously separations are large but acyclical. We build our alternative, no-selection parameterization so as to have identical steady-state levels of output and unemployment but larger average labor flows, relative to the economy with cyclical selection.

The parameterization for the no-selection economy is given in Table 3. It reduces the average productivity advantage of high-efficiency workers from 18% to 1%. However, low-efficiency workers have a large productivity variance when employed in a match, with hourly output ranging from a minimum of 0% to a maximum of about 200% of high-efficiency workers hourly output. This results in the steady state unemployment share of low-efficiency workers rising from 30% in the model with selection to 53% in the no-selection model, while the endogenous separation rate rises from 3.1% to 20%. To ensure large enough unemployment-to-employment flows corresponding to the much higher employment-to-unemployment flow, the parameterization requires a higher matching efficiency, and a lower vacancy posting cost.

5.1 The COVID recession

The first result of the model is the quantification of the shocks necessary to match the given path for total separations. The upper left panel of figure 4 shows the paths for the shock to ρ^x in the selection (solid lines) and no-selection (dashed lines) models. Given endogenous separations are acyclical in the no-selection economy, it requires a much larger exogenous separation rate shock to match the same total separation rate as in our benchmark economy with selection. Thus, ρ_t^x increases by 150% in the economy without selection, relative to its steady state value. This represents a rise of ρ_t^x from 6.8% to 17%. In the model with selection, endogenous separations move significantly (see upper right panel). With selection, the endogenous separation rate for type l workers increases from 3.1% to 27.9% per quarter. Since our focus is on the different impact of random and selected separations, we assume the

[14] Large steady state labor flows were shown in Ravenna and Walsh (2012b) to virtually eliminate selection in the face of productivity shocks and to generate virtually constant shares of low-efficiency workers in the pool of unemployed over the business cycle.

demand shock is identical in both the selection and no-selection economies, and is obtained from jointly matching the path of separations and output in the economy with selection. This lets the simulation isolate the impact of selection on both employment and output.

The second row of the figure shows output and the overall unemployment rate. Selection amplifies the peak fall in output from 8.2% to 10.5%. This is the consequence of the much larger increase in unemployment when selection is at work, pushing the joblessness rate to over 10 percentage points above steady state, to 15.7%. Selection leads to an increase in the low-efficiency unemployment share of 134% relative to pre-recession levels, equivalent to 40 log-points (third row, left panel). As a consequence, the difference between the productivity of the average employed and the average unemployed worker rises as the COVID recession persists, with the expected productivity of a new hire falling by 6% (shown in third row, right panel), even though the exogenous level of TFP in the economy's is unchanged.

An environment with a higher exogenous separation rate, as in the no-selection model, lowers the future value of employment for job seekers and for firms. The lower right panel of figure 4 shows that this produces a large negative effect on vacancies per unemployed worker. In the model with selection, a larger share of separations arise from endogenous separations.[15] Rather than generating a similarly sized fall in employment across all workers, selection results in a much higher rate of endogenous separations and therefore a differential impact on low-efficiency workers. A positive shock to ρ^x lowers the surplus for every worker proportionally (by lowering continuation values), and since the expected surplus is always lower for l-workers, it requires a higher match-specific productivity level for an l-worker match to generate a positive surplus. This implies $\bar{a}_t$ rises, and firms become more selective. Thus, even if the shock to the exogenous separation rate initially affects all workers regardless of type, selection amplifies the impact on low-efficient workers.

The model delivers a 'paradox of separations': random separations, affecting workers of both h and l type, are inefficient from the point of view of the firm and the worker as they destroy matches that were generating a positive surplus. However, random separations do not generate a large impact on unemployment per unit of output lost. They also do not generate a large inefficiency in the allocation, since they have a limited impact on average productivity

[15] Recall, total separations are the same in both models. See the top row of figure 4.

among the unemployed. On the other hand, endogenous separations, which are optimally chosen by firms, result in a large increase in unemployment relative to the decline in output, and, by affecting the average quality of the pool of job seekers, imply a larger inefficiency in the allocation.

In the no-selection economy, the unemployment share of l-workers falls, as the increase in exogenous separations affecting both worker types is much larger. However average productivity among unemployed workers barely moves, as h- and l-workers are on average very similar in terms of productivity, and the COVID-induced change in unemployment shares across worker-types is modest relative to the steady state. As a consequence, most of the fall in output is driven by a persistent fall in hours per worker (see row four). In line with the experience of the U.S. in the Great Recession, hours per worker in the selection economy rebound much faster than unemployment; hours are back to their pre-recession level when unemployment has fallen by only one quarter of the total rise following the COVID shock.

The COVID recession has a very different impact across subgroups of workers. The top row of figure 5 shows the impact on unemployment rate of low (on left) and high (on right) productivity workers. Low-efficiency workers experience the largest rise in joblessness, both in absolute and in percentage terms. Compared to an economy without selection, the endogenous increase in separation strongly biases the impact of the COVID shock towards the group of workers who have on average the worse labor market outcomes. Selection has virtually no impact on unemployment for h-workers, whose unemployment increases under both parameterization by less than 2.5 percentage points above its steady state value. In sharp contrast, l-workers see their unemployment rate rise by over 35 percentage points. While selection amplifies the COVID-recession impact on l-workers, even in the absence of selection they experience a much larger rise in unemployment compared to h-workers.

This large impact in the baseline model is the consequence of selection occurring both at separation and at hiring. The latter is key to explaining the high and persistent rate of joblessness among the workers belonging to the low-productivity group, and eventually for the labor market dynamics of the aggregate economy. The lower left panel of figure 5 shows that the fall in the job-finding probability during our COVID scenario in the selection economy is more than three times larger for l-workers. This contributes to the fall in the share of l-workers in the employment pool by 6 percentage points, or 25% of its steady state value (lower right panel).

The earlier discussion clarified that $l-$workers bear the brunt of the increase in joblessness and unemployment duration during the COVID recession. The impact on this group of workers is driven principally by an unobservable variable – the drop in average productivity among the job seekers – that in the economy cannot be easily inferred from observable variables related to the employed pool of workers. Surprisingly, selection results in a marked disconnect between labor productivity, unemployment and wages, shown in figures 6 and 7. Total earnings per worker fall among h-workers (see figure 6), and follow a comparable dynamics over the recovery, regardless of whether selection is at work or not. In contrast, selection leads to an increase in total earnings for the set of low-efficiency workers who manage to be employed. This is the result of firms becoming much more selective and hiring only the best among the l-workers. Endogenously we observe wage compression among the employed.

Average earnings across all employed workers fall initially because of the fall in hours per worker and of the exogenous increase in the separation rate, lowering the surplus for h-workers, and for the average l-worker. Over time, as firms replace l-workers with h-workers, and as those l-workers who are employed have higher productivity relative to pre-recession levels, the wage gap caused by the COVID shock is quickly closed, both in hourly and in total earnings. The left panel of figure 7 shows that the wage gap is closed far earlier than the unemployment gap. Selection makes wages endogenously sticky, despite our assumption that wages are determined by Nash bargaining in every period. Wage growth is zero by the fifth quarter, while aggregate unemployment is still 5% above its pre-recession level.[16]

The behavior of hourly wages depend on the endogenous movements of the total surplus generated by a match and the number of hours supplied. Match surplus falls during the downturn, but part of the fall is explained by the endogenous fall in hours worked per match. Thus, in the model hourly wages do not fall as much as does the total surplus. At the same time, this does not prevent the fall in match surplus from having a strong impact on hiring. As the total surplus falls, ceteris paribus, the implied cost of a match increases since firms equate

[16] During the Great Recession and the subsequent recovery, compensation growth did not slow as much as expected, given the amount of slack in the economy and the high level of unemployment experienced after the initial downturn (Linder, Peach and Rich, 2012).

the fixed vacancy cost with the total, rather than the hourly, surplus.

The right panel of figure 7 shows that positive selection causes average productivity among the employed to rise during the COVID recession.[17] On impact, the COVID shock leads simultaneously to a lower wage and a higher labor productivity. During the recovery, wage inflation is positive or zero, while labor productivity growth is negative. The reason for this disconnect is that wages reflect the expected value of the match over its lifetime, while labor productivity only reflects the current output of the match.

6 The impact of monetary policy and the ZLB

The equilibrium responses of the economy in a new Keynesian model depend on the specification of monetary policy, and the results reported so far have been based on a non-inertial instrument rule in which the nominal interest rate responds only to inflation. And, importantly, the zero lower bound constraint has been ignored. In this section, we examine the role of the composition effect on labor market outcomes when monetary policy is limited by the ZLB.

It is useful, however, to first review how monetary policy affects output, employment and inflation in this model. A reduction in the real interest rate, operating through the household's Euler condition, leads households to increase current consumption. This increases the demand for final goods, and retail firms respond by increasing production as in standard new Keynesian models. To increase production, retail firms must purchase more wholesale goods, and because retail prices are sticky, this pushes up the price of wholesale goods relative to retail goods; the retail price markup falls, generating a countercyclical movement in the markup.[18]

The movement in the markup affects several margins relevant for separations, job creation and hours.[19] Consider first the effects on high-efficiency workers operating through the markup.

[17] The fall in the cyclicality of labor productivity over the last three recessions has been extensively discussed in the literature (Berger, 2012, Gali and van Rens, 2010). Petrosky-Nadeau (2012) and Fernald (2013) find that the dynamics of labor productivity over the Great Recession have not deviated much from its trend.

[18] Nekarda and Ramey (2013) argue that markups are procyclical in the U.S.. However, this is not inconsistent with the effects of monetary policy discussed here, as productivity shocks would generate a procyclical markup.

[19] As shown by Andrés et al. (2012), the presence of price rigidities, even absent wage rigidities, contributes significantly to the ability of the basic search and matching model to match the volatility of labor market tightness and vacancies, as well as generate a negatively sloped Beveridge curve. And Lago Alves (2014) shows that introducing a non-zero trend rate of inflation when prices adjust according to the Calvo model increases the volatility of unemployment sufficiently to solve the Shimer puzzle even if all wages are flexible.

The surplus associated with such workers, given by (5), rises when μ_t falls. This increases the incentive for wholesale firms to post vacancies. From (6), the fall in μ_t also increases the optimal hours worked by high-efficiency workers. So wholesale firms increase hours of existing high-efficiency employees and post more job vacancies.

The surplus generated by a low-efficiency worker is similarly increased by a fall in the markup, as are their optimal hours (see 7 and 8). However, the markup also affects the cutoff productivity level $\bar{a}_t$ given by (9). A fall in μ_t lowers $\bar{a}_t$, implying a fall in endogenous separations and a fall in the screening rate – more low-efficiency unemployed workers who are interviewed are subsequently hired (see 2). Thus, hours rise, vacancies rise, fewer low-efficiency workers are screened out, and match efficiency rises.

In addition, current match surpluses and the critical cutoff productivity level determining endogenous separations and low-efficiency hiring also depend on the continuation value of a match. An expansionary monetary policy shock lowers the real interest rate, and this increases the present discounted value of future surpluses.[20] This increases job creation, reduces endogenous separations, and leads more low-efficiency unemployed workers to be hired at the interview stage. While the present model abstracts from physical capital, a fall in the real interest rate leads wholesale firms to increase their investment in matches.

A final channel of monetary policy arises because job posting costs generate a demand for final goods on the part of wholesale firms. Expansionary monetary policy, by stimulating job creation, leads to a direct increase in the demand for final goods.

6.1 The COVID recession and the ZLB

We explore the role the zero lower bound on nominal interest rates plays in limiting the ability of monetary policy to help engineer a more rapid recovery from the COVID recession. Several OECD countries have been experimenting with negative interest rates, an option which at the moment the Federal Reserve has not included in its set of non-conventional policies.

To solve for the law of motion at the zero lower bound for i_t, we adapt the code in Guerrieri and Iacoviello (2015). In our model the future behavior of the shocks leading the economy to the zero lower bound is fully expected as of the first period when the shocks arrive. The

[20] See Hall (2017).

solution method solves models with occasionally binding constraints by building the time-varying decision rule for the whole period when the constraint binds.[21]

Figure 8 presents the COVID recession for the baseline selection model under the same set of shocks discussed in the earlier section, comparing the outcome to an economy in which the zero-lower bound is binding.[22] The solid line shows the responses in the baseline model when the ZLB constrains monetary policy. The top two panels and the left panel of row 2 show that worker heterogeneity has little effect on the severity or persistence of the effects on output and inflation after the first four quarters. However, the ZLB constraint greatly amplifies the recession in the first year after the COVID shock and causes unemployment to rise by a further 6 percentage points. Figure 4 showed that time-varying selection among workers amplifies the effect of the COVID shock, and the constraint on monetary policy further exacerbates the adverse impact of the COVID shock. In the medium run, the selection effect is the primary driver of the slow employment and output recovery. Overall, having only a limited policy space for the monetary authority is especially detrimental once we account for the strong impact on unemployment resulting from selection among workers.

Figure 9 shows that the cost of the lack of policy space is disproportionately bore by low-efficiency workers. Their unemployment rate increases by about 50% more at its peak – or 20 additional log-points – because of the ZLB, while the unemployment rate increase for high-efficiency workers only rises from 2.5% to 3.8% at its peak. However, increased selection among employed workers leads to no change in average earnings among l-workers (whose average quality improves further when the ZLB constraint is binding), while the fall in earnings for h-workers more than doubles, as a consequence of the deeper recession.

While the COVID shock is considered at least in part a supply shock, the fall in total industry U.S. capacity utilization from 78% in February to 65% in April has lowered fears that the shock would prove inflationary. Accounting for the ZLB constraint on monetary policy is

[21]The algorithm starts with a guess as for the duration of the period where the constraint binds, building expectations using the recursive law of motion for the linear approximation to the model when the constraint does not bind. Moving backward, it can then build the law of motion at each point in time when the constraint is binding. The algorithm iterates until no violation of the constraint remains. This solution method allows for endogenously determining the horizon over which the ZLB is a binding constraint, given the dynamics of the shock.

[22]To drive the economy to the ZLB, we introduce an additional persistent shock lowering the discount rate by approximately one third of its steady state value. This shock implies in equilibium a negligible impact on real variables, and a persistent fall in nominal variables, that we abstract from in the figures.

predicted by our model (see figure 8) to lead to a more deflationary outcome under the COVID scenario, with inflation up to 8 percentage points below its trend value in annualized terms. However, inflation rapidly converges to slightly below trend after four quarters. We take this prediction with some caution, as inflation in the model is completely forward looking, and wages are allowed to change flexibly in every quarter.

7 Temporary layoffs

Our baseline model, in common with standard DMP models, assumes that all exogenous separations are permanent separations. Workers flowing into unemployment directly enter the pool of searching workers.[23] As Barrero et al. (2020) and Kudlyak and Wolcott (2020) have emphasized, a large fraction of layoffs due to the COVID-19 pandemic have taken the form of temporary layoffs.

According to the BLS, the separation rate for the private sector, including both layoffs and quits, increased from a monthly rate of 4.1% to 11.1% in March 2020 and 8.8% in April 2020. Within the excess-inflow into unemployment since the start of the COVID recession, the BLS household survey finds the share of workers on temporary layoffs in total unemployment, which averaging 13.5% over the period from January 1985 to March 2020, averaged over 75% during April and May 2020 before falling to 59.5% in June 2020. Based on previous recessions, Barrero et al. (2020) estimate that 42% of recent layoffs will result in permanent job losses. It is therefore important to see how dynamics are affected when a large share of layoffs are temporary. If these workers are recalled to their former jobs as the economy begins to re-open, the speed at which employment will recover may be much faster than predicted by models that ignore temporary layoffs.[24] In this section, therefore, we extend our model to include both permanent and temporary layoffs.

[23] Because we allow workers who do not survive the exogenous separation hazard to immediately search for a new match, some will quickly find a new match and so could be considered similar to those on a temporary layoff. For type h workers, the fraction k_t^w will be rehired. However, because k_t^w falls in recessions, the share of recent exogenous separations that move immediately to a new match is negatively correlated with the level of unemployment.

[24] Some countries have indeed observed fast rebounds: Norway saw its unemployment rate climb from 2% to 11% within 60 days since its first COVID cases, and then a fall back to 7% after a further 30 days as the economy started reopening. However, many separations are expected to become permanent, as future pandemic prevention policies may call for structural reductions in some sectors, such as travel and hospitality.

It is useful to note first how unusual the COVID-19 behavior of temporary layoffs is. Over the 423 months between January 1985 and March 2020, a period that includes the Great Moderation, the Global Financial Crisis, and the post-financial-crisis recovery, the covariance between the unemployment share of temporary layoffs and the unemployment rate was -0.58, indicating that the share of workers on temporary layoffs was procyclical. Adding the three months April to June 2020 to the sample causes this covariance to flip to a positive 0.15. The COVID surge in both unemployment and in the share of those on temporary layoffs is unprecedented.

To capture this phenomenon, we extend the model and assume a fraction $0 < \delta_{T,t} < 1$ of all exogenous separations are temporary. If δ_T were constant, though, our model would predict that the share of the unemployed on temporary layoff would fall in a recession, consistent with the pre-pandemic evidence, as total separation include both exogenous and endogenous separations and the later rise during a recessions. To capture the positive co-movement seen during the pandemic, we assume the positive shock to total exogenous separations is accompanied by a positive shock to $\delta_{T,t}$. We then assume workers on temporary layoff are recalled at a constant rate.

Consistent with our basic model which set the productivity of type h workers such that they always generate a positive match surplus, we assume any type h worker recalled from temporary layoff is hired. However, a type l worker who is recalled from temporary layoff may no longer generate a positive surplus at the time of recall, both because their idiosyncratic productivity realization has changed and because the critical cutoff value of productivity will have changed. Thus, at time $t+i$ only the fraction $\left(1 - \rho^n_{t+i}\right)$ of recalled workers of type l will survive screening and actually be hired. We assume those who are screened out after recall enter the pool of permanently separated job seekers.

7.1 Calibration

To assess the effects of temporary layoffs on the dynamic adjustments to our COVID recession scenario, we utilize the same exogenous separations and demand shock realizations used in section 5. The scenario discussed in section 5 assumed $\delta_{T,t} \equiv 0$, but we now allow for $\delta_{T,t} > 0$ so that a fraction of the exogenous separations represent temporary layoffs.

To allow the introduction of temporary separations to have the strongest possible impact on the dynamics of the model, we assume that all workers flowing into the pool of temporary layoffs eventually get recalled for a job interview without the need of a vacancy being posted, and that the probability of a recall interview within four quarters of the initial separation is 95%. This implies a quarterly-recall hazard of 53%. In steady state, over 99% of recalled workers enters into a productive match.

We parameterize the steady state share of workers on temporary unemployment relative to the total stock of unemployed to 13%, a value in line with the share reported in Kudlyak and Wolcott (2020) for the 1985-2019 sample. This share, together with the quarterly recall rate, implies that the steady state share of exogenous separations flowing into the pool of temporary unemployment is equal to 6.44%.

The shock $\delta_{T,t}$ is set so that the share of workers on temporary unemployment relative to the total stock of unemployed in the first quarter is equal to 70%, approximately equal to the average of the temporary unemployment share over the months of March, April, May 2020. We assume that the temporary unemployment share declines in the second, third and fourth quarter to 60%, 32% and 9%, implying that within a year the ratio falls below its steady state value, as the pool of temporary unemployed is reabsorbed faster than the pool of permanently separated unemployed who need to be matched to a new vacancy.

7.2 Pandemic scenario with temporary layoffs and recalls

In the COVID-scenario with temporary layoffs the initial increase in exogenous separations, while falling equally on both worker type, does not lead to an increase in the pool of job seekers that is as large as under the earlier scenario as those on temporary layoff await recall. The positive shock to $\delta_{T,t}$ captures the positive co-movement of unemployment and the share on temporary layoff seen in the COVID recession.

Figure 10 compares responses in the baseline model with only permanent layoffs (solid lines) to the model with both permanent and temporary layoffs (dashed lines). The top row of the figure shows that the presence of temporary layoffs initially leads to a larger fall in output and rise in unemployment, but the maximum fall in output (rise in unemployment) is slightly smaller, and after three quarters, the unemployment rate is lower then in the permanent layoff case. The source of the larger initial rise in unemployment reflects the timing in our model; a worker on permanent unemployment immediately searches for work, and some fraction of these workers will obtain jobs. Workers on temporary layoff wait to be recalled. Once recalls begin, unemployment declines more quickly than in the baseline scenario.

The middle left panel of figure 10 shows that the unemployment share accounted for by low-productivity workers is even higher when a significant fraction of layoffs are temporary, and their unemployment rate is higher, as seen in the upper left panel of figure 11. The reason for these adverse effects is that type-l workers who are recalled are only hired if they will generate a positive surplus, and this requires that their individual productivity exceed the cutoff value of $\bar{a}_t$ at the time they are recalled. Those workers who are recalled but not rehired must then enter the pool of job seekers. In fact, employment prospects for type l workers are worsened by the use of temporary layoffs through three channels. First, type l workers that experience a temporary layoff do not immediately search for work as they await recall. This results in higher unemployment, as it also does for type h workers. Second, the COVID-induced recession reduces match surplus. This increases the critical value of $\bar{a}_t$ that governs endogenous separations, so in addition to the rise in exogenous separations, type l workers experience a rise in endogenous separations. And third, when type l workers are recalled, they are only rehired if they generate a positive surplus; given the rise in $\bar{a}_t$, fewer such workers are actually hired when recalled.

This contrasts with the unemployment experience of type-h workers, whose unemployment rate is shown in the top right panel of figure 11. Their unemployment rate initially is higher since those put on temporary layoff do not immediately becomes job seekers, but all high-productivity workers are rehired when recalled. Thus, as firms recall workers, unemployment for type-h workers declines rapidly, particularly in comparison to the baseline case (solid line) with only permanent layoffs. For high-productivity workers, temporary layoffs do lead to a V-shaped employment recovery. The top right panel of figure 10 shows that the overall unemployment rate inherits the slow recovery that characterizes unemployment among low-productivity workers, and this is true with temporary layoffs (dashed lines) or with only permanent layoffs (solid line).

The shock to the exogenous separation rate, a mass layoff, affects both types of workers and leads to a large inflow into unemployment of both type l workers and type h workers. The induced increase in selectivity and rise in endogenous separations means the pool of employed workers is now more heavily weighed toward high-quality workers, as can be seen in the lower right panel of figure 11. The shift is similar whether all layoffs are permanent or a large fraction are temporary. The lower right panel shows the share of workers on temporary layoff as a percent of total unemployment. It peaks at 70% of total unemployment and then falls as those on temporary layoff are recalled. After four periods, the share has fallen to 10% of the stock of unemployment.

Mass layoffs of which a large fraction are temporary do dampen the peak rise in the unemployment rate and speeds its decline slightly. However the main effect of temporary layoffs is to accentuate the differences in unemployment across worker types.

8 Conclusions

We use a simple model of labor heterogeneity in a new Keynesian search and matching framework to show how macro shocks have differential effects across workers because those workers who experience separations and those job seekers who are hired are not drawn randomly from the relevant population. Instead, a process of selection is at work. We highlight how selection amplifies the overall unemployment consequences of contractionary shocks such as.aggregate demand and labor market shocks of the type associated with the COVID recession. In our basic framework, workers who, on average, experience more frequent and longer spells of unemployment also experience the largest decline in employment in recessions.

Selection in separations and in hiring can lead to a disconnect between the average productivity of employed workers and the incentive to hire unemployed workers. In a recession, the composition of the employed shifts toward high-efficiency workers and toward more productivity low-efficiency workers. The composition of job seekers shifts toward low-efficiency workers. Measured productivity of the employed can rise while average productivity of job seekers falls, and it is the latter that drives the incentive to post vacancy and form new matches. A fall in the incentive to post vacancies reduces the job-finding rate for all job seekers, and leads to a persistent increase in unemployment.

We show how heterogeneity can generate an externality; firms take the composition of the pool of unemployed job seekers as given, yet their decisions to end matches or hire workers affects the composition of that pool. This externality, and the resulting distortion it generates, are present even under Nash bargaining with the Hosios condition satisfied, price stability is maintained, and the steady-state distortion due to imperfect competition is eliminated. Thus, the adverse impact of recessions on low-productivity workers that arises from selection is socially costly. In our framework, the effects of the COVID recession are also significantly worsened when monetary policy's ability to respond is constrained by the zero lower bound on the nominal interest rate.

At the time we are writing, the eventual impact of the COVID pandemic on the economy is unknown. Our baseline model with selection suggests the recovery will be L-shaped. However, this baseline treated all separations (exogenous or endogenous) as permanent. The recent experience is that a large fraction of layoffs are reported as temporary, and this suggests it may be possible to have a V-shaped recovery. We extend our model to allow for temporary layoffs. We find that they do speed the employment recovery for high-productivity workers; their unemployment rate falls quickly back to its steady-state value. However, this is not the case for low-productivity workers on temporary layoff who experience a long, slow employment recovery. And the overall recovery of unemployment is only slightly improved when most separations take the form of temporary layoffs.

Our major focus has been on investigating the role of labor market selection on macro dynamics. To do so, we have employed a one-sector model. Guerrieri et al. (2020) has shown how negative supply shocks, such as the separation shocks we consider, can have Keynesian demand effects in a multi-sector model. We have also ignored variations in labor force participation. The impact of COVID has causes a sharp drop in labor force participation (LFP) in the U.S. At this point, it is hard to know whether this is simply a discouraged worker effect, as workers recognize that with many firms in lockdown, few are hiring. In fact, this scenario would suggest many of those who have left the labor force will return as the economy recovers. However, the CARES Act passed by Congress, which provides increased benefits for the unemployed, would provide incentives to remain in the labor force. This might suggest those leaving are not likely to return quickly. We note that the large drop in the LFP rate relative to the rise in the unemployment rate is roughly the same as occurred during the Great Recession. The behavior of temporary layoffs in the COVID recession is, in contrast, quite different, and this is a major reason for focusing on it in this paper. Incorporating a labor force participation market and investigating selection in the context of a multi-sector model are interesting extensions to the model we have left for future work.

References

Ahn, H. J. and J. D. Hamilton (2019). Heterogeneity and Unemployment Dynamics. *Journal of Business & Economic Statistics*, 1–58.

Andrés, J., R. Domenech, and J. Ferri (2012). Price Rigidity and the Volatility of Vacancies and Unemployment. *Working Paper, nternational Economics Institute, University of Valencia*.

Baker, G. P. (1992). Incentive Contracts and Performance Measurement. *Journal of Political Economy 100*(3), 598–614.

Baley, I., A. Figueiredo, and R. Ulbricht (2020). Mismatch Cycles. *SSRN Electronic Journal*.

Barrero, J. M., N. Bloom, and S. J. Davis (2020). COVID-19 is also a reallocation shock. *NBER Working Paper No. 27137*.

Berger, D., I. Dew-Becker, L. D. W. Schmidt, and Y. Takahashi (2019). Layoff Risk, the Welfare Cost of Business Cycles, and Monetary Policy. *Working Paper*.

Bick, A. and A. Blandin (2020). Real Time Labor Market Estimates During the 2020 Coronavirus Outbreak. *Manuscript, Arizona State University*.

Bils, M., Y. Chang, and S.-B. Kim (2012). Comparative advantage and unemployment. *Journal of Monetary Economics 59*(2), 150–165.

Blanchard, O. and J. Galí (2007). Real Wage Rigidities and the New Keynesian Model. *Journal of Money, Credit and Banking 39*(1), 35–65.

Blanchard, O. and J. Galí (2010). Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics 2*(2), 1–30.

Christiano, L., M. Eichenbaum, and M. Trabandt (2016). Unemployment and Business Cycles. *Econometrica 84*, 1523–1569.

Davis, S. J., J. C. Haltiwanger, and S. Schuh (1996). *Job Creation and Job Destruction.* Cambridge, M. A: The MIT Press.

Dickens, W. T. and R. K. Triest (2012, jan). Potential Effects of the Great Recession on the U.S. Labor Market. *The B.E. Journal of Macroeconomics 12*(3).

Fujita, S. and G. Ramey (2009, may). The Cyclicality of Separation and Job Finding Rates. *International Economic Review 50*(2), 415–430.

Galí, J. (2011). *Unemployment Fluctuations and Stabilization Policies: A New Keynesian Perspective.* Cambridge, M. A: The MIT Press.

Gertler, M., L. Sala, and A. Trigari (2008). An estimated monetary DSGE model with unemployment and staggered Nash wage bargaining. *Journal of Money, Credit and Banking 40*(8), 1713–1764.

Gertler, M. and A. Trigari (2009). Unemployment Fluctuations with Staggered Nash Wage Bargaining. *Journal of Political Economy 117*(1), 38–86.

Gregory, V., G. Menzio, and D. G. Wiczer (2020). Pandemic Recession: L or V-shaped? *NBER Working Paper No. 27105*.

Grigsby, J. (2020). Skill Heterogeneity and Aggregate Labor Market. *Working paper*, 0–83.

Guerrieri, L. and M. Iacoviello (2015). OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics 70*, 22–38.

Guerrieri, V. (2007). Heterogeneity and unemployment volatility. *Scandinavian Journal of Economics 109*(4), 6667–693.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning (2020). Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages? *NBER Working Paper No. 26918*.

Hall, R. E. (2015). Quantifying the Lasting Harm to the U. S. Economy from the Financial Crisis. In J. A. Parker and M. Woodford (Eds.), *NBER Macroeconomics Annual 2014*, Volume 29, pp. 71–128. University of Chicago Press.

Hall, R. E. (2017). High discounts and high unemployment. *American Economic Review 107*(2), 305–330.

Hall, R. E. and S. Schulhofer-Wohl (2018). Measuring job-finding rates and matching efficiency with heterogeneous job-seekers. *American Economic Journal: Macroeconomics 10*(1), 1–32.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review 101*(7), 2873–2898.

Kahn, L. B., F. Lange, and D. G. Wiczer (2020). Labor Demand in the Time of COVID-19: Evidence from Vacancy Postings and UI Claims. *NBER Working Papers*.

Kapicke, M. and P. Rupert (2020). Labor Markets during Pandemics.

Kudlyak, M. and E. Wolcott (2020). Pandemic Layoffs. (May), 1–21.

Lago Alves, S. A. (2014). Lack of Divine Coincidence in New Keynesian Models. *Journal of Monetary Economics 67*, 33–46.

Leduc, S. and Z. Liu (2020). The weak job recovery in a macro model of search and recruiting intensity. *American Economic Journal: Macroeconomics 12*(1), 310–343.

Lemieux, T. (2006). Increasing residual wage inequality: composition effects, noisy data or rising demand for skills? *American Economic Review 96*(3), 461–498.

Nagypal, E. (2007). Learning-by-doing versus learning about match quality. *Review of Economic Studies 74*(2), 537–566.

Nagypal, E. and D. T. Mortensen (2007). Labor-market volatility in matching models with endogenous separations. *Scandinavian Journal of Economics 109*(4), 645–665.

Nekarda, C. and V. Ramey (2013). The Cyclical Behavior of the Price-Cost Markup. *NBER Macroeconomic Annual*.

Pries, M. J. (2008). Worker heterogeneity and labor market volatility in matching models. *Review of Economic Dynamics2 11*, 644–687.

Pries, M. J. and R. Rogerson (2005). Hiring policies, labor market institutions, and labor market flows. *Journal of Political Economy 113*(4).

Ravenna, F. and C. E. Walsh (2008). Vacancies, unemployment, and the Phillips curve. *European Economic Review 52*(8).

Ravenna, F. and C. E. Walsh (2011). Welfare-Based Optimal Monetary Policy with Unemployment and Sticky Prices: a Linear-Quadratic Framework. *American Economic Journal: Macroeconomics 3*(April), 130–162.

Ravenna, F. and C. E. Walsh (2012a, mar). Monetary Policy and Labor Market Frictions: A Tax Interpretation. *Journal of Monetary Economics 59*(2), 180–195.

Ravenna, F. and C. E. Walsh (2012b, mar). Screening and Labor Market Flows in a Model with Heterogeneous Workers. *Journal of Money, Credit and Banking 44*(2), 31–71.

Ravn, M. O. and V. Sterk (2017). Job uncertainty and deep recessions. *Journal of Monetary Economics 90*, 125–141.

Trigari, A. (2009, feb). Equilibrium Unemployment, Job Flows, and Inflation Dynamics. *Journal of Money, Credit and Banking 41*(1), 1–33.

Walsh, C. E. (2003). Labor Market Search and Monetary Shocks. In S. AltuÄİ, J. Chadha, and C. Nolan (Eds.), *Elements of Dynamic Macroeconomic Analysis*, pp. 451–486. Cambridge, U.K.: Cambridge Univeristy Press.

Table 1: Baseline Parameterization		
Steady State Values		
Unemployment rate	u_{ss}	5.7%
Unemployment rate - $l-efficiency$ labor	u^l_{ss}	11.6%
Unemployment rate - $h-efficiency$ labor	u^h_{ss}	4.4%
Average hours per worker	h^{av}_{ss}	0.33
Vacancy posting cost share of output	$\frac{\kappa V_{ss}}{Y_{ss}}$	0.015
Probability of vacancy matched with applicant	k^f_{ss}	0.9
Parameters		
Unemployment elasticity of matching function	α	0.6
Discount factor	β	0.99
Inverse of labor hours supply elasticity	χ	1
Relative risk aversion	σ	1
Gross steady state inflation rate	π_{ss}	1
Workers' share of surplus	η	0.6
Steady state separation rate	ρ_{ss}	7.45%
Exogenous separation rate	ρ^x	6.8%
Endogenous separation/screening rate	ρ^n_{ss}	3.1%
AR(1) parameter for exogenous shocks	ρ_z	0.95
Price elasticity of retail goods demand	ε	6
Average retail price duration (quarters)	$\frac{1}{1-\omega}$	4

Table 1: Baseline Parameterization. US unemployment rate for low- and high-efficiency workers is given respectively by the rate for the 16 to 24 and over-24 age-group, over the 1948:1-2019 sample (BLS, 2019:4).

Table 2: Implied steady state productivity and labor market variables			
			Baseline
Productivity	Average productivity of high-efficiency workers		0.591
	Average productivity of low-efficiency workers		0.50
	Relative productivity of high/low-efficiency workers		1.18
	Average productivity labor force		0.57
	Average productivity employed workers		0.603
	Average productivity unemployed workers		0.564
Steady State	low-efficiency unemployment share	γ_{ss}	0.30
	low-efficiency employment share	ξ_{ss}	0.22
	low-efficiency labor force share	$\overline{\gamma}$	0.233
	Unemployment duration (quarters)		1.71

Table 2: Implied steady state productivity and labor market variables. Average productivity of high and low-efficiency worker-hours is given by ϕ^h and $\phi^l \int_0^1 a^l_i dF(a^l_i)$.

Table 3: Alternative Parameterization: Selection vs. No-Selection Economy

			Selection	No Selection
Productivity	Average productivity of high-efficiency workers		0.591	0.562
	Average productivity of low-efficiency workers		0.50	0.56
	Relative productivity of high/low-efficiency workers		1.18	1.01
	Average productivity labor force		0.57	0.561
	Average productivity employed workers		0.575	0.583
	Average productivity unemployed workers		0.603	0.56
Parameters	Matching function productivity	ψ	0.70	0.84
	Vacancy posting cost	κ	0.08	0.025
Steady State	Overall separation rate	ρ_{ss}	7.4%	10%
	Endogenous separation/screening rate	ρ^n_{ss}	3.1%	20%
	low-efficiency unemployment share	γ_{ss}	0.30	0.53
	low-efficiency employment share	ξ_{ss}	0.22	0.19
	low-efficiency labor force share	$\overline{\gamma}$	0.233	
	Unemployment duration (quarters)		1.71	1.55

Table 3: Alternative Parameterization: Selection vs. No-Selection Economy. Average productivity of high and low-efficiency worker-hours is given by ϕ^h and $\phi^l \int_0^1 a_i^l dF(a_i^l)$. The two parameterization have identical steady state output and unemployment

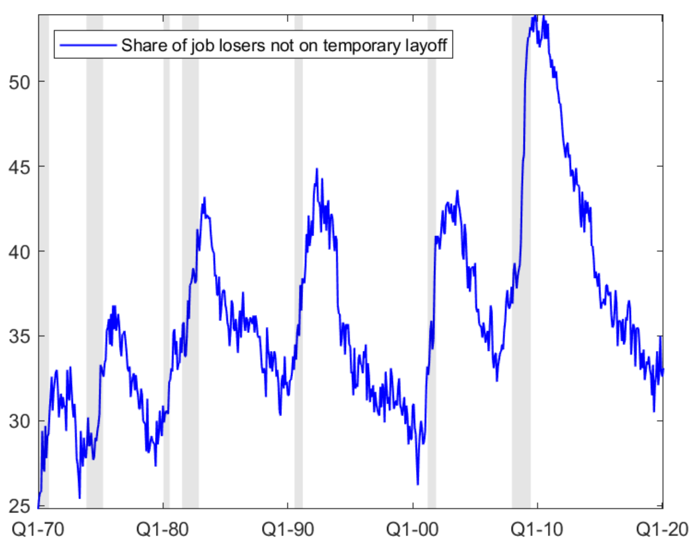

Figure 1: Percent of total U.S. unemployed 16 years old and over classified as job losers not on temporary layoff as reason for unemployment, Jan. 1967 - Mar. 2020. Seasonally adjusted monthly data. Shaded areas indicate NBER recession dates. Source: CPS series LNS13026511.

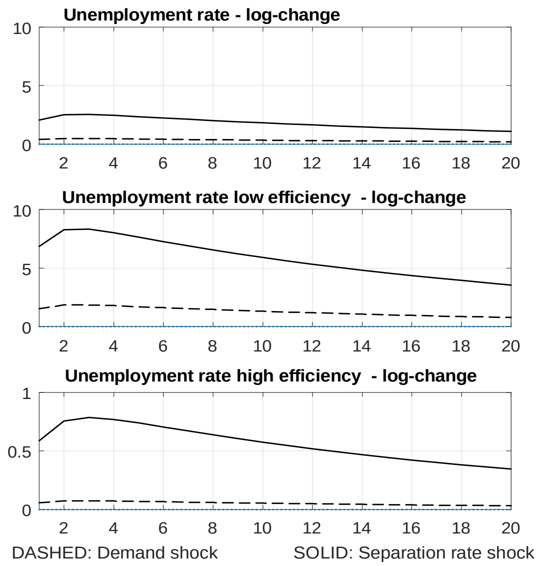

Figure 2: Response of aggregate unemployment and unemployment for h and l workers to a demand shock (dashed lines) and to a separation shock (solid lines). Both shocks parameterized to generate a 1% fall in output on impact. (Note different scale used for lower panel.)

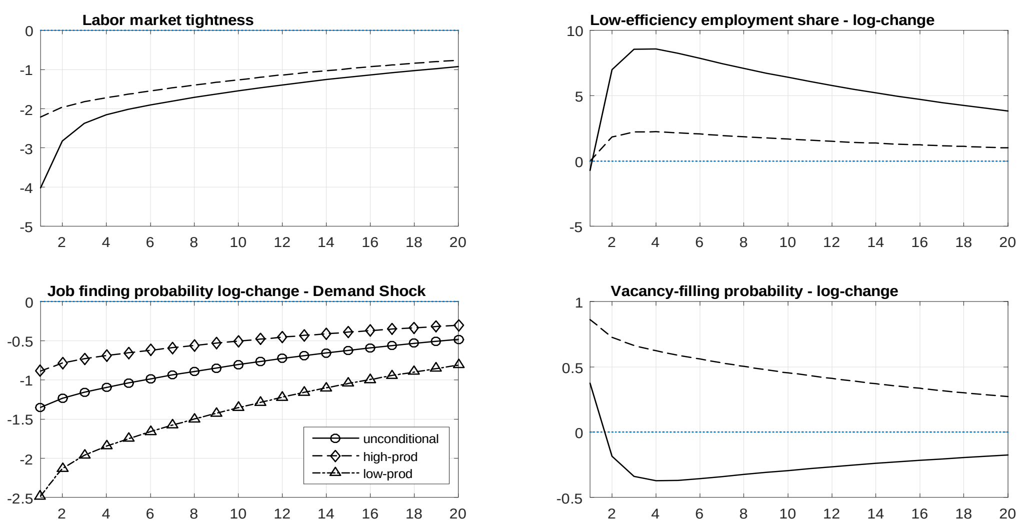

Figure 3: Response to a demand shock and a separation shock in selection model. Both shocks parameterized to generate a 1% fall in output on impact. In panels 1,2 and 4: solid line indicates separation shock, dashed line indicates demand shock. Panel 3: job finding probability conditional on demand shock.

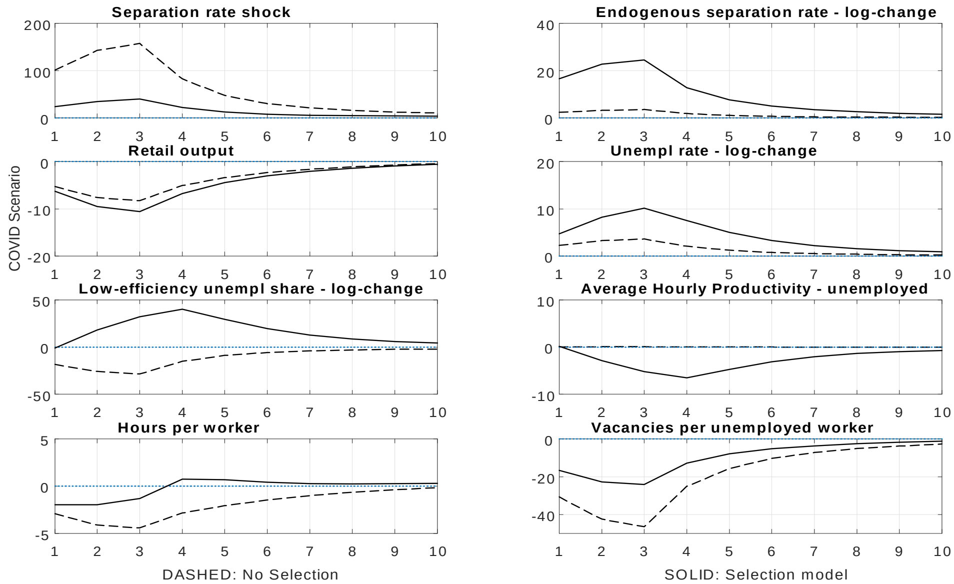

Figure 4: COVID scenario - selection model (solid lines) and selection model (dashed lines). Demand shock and separation shock parameterized to match given path of output and of total separations in the selection and no-selection models. Demand shock is assumed identical across the two models.

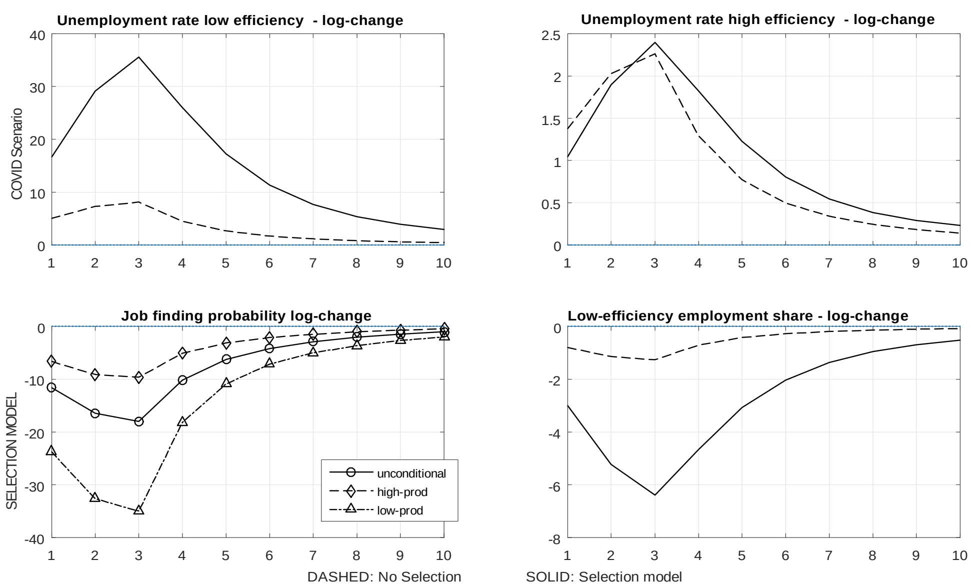

Figure 5: COVID scenario - selection model (solid lines) and selection model (dashed lines). Demand shock and separation shock parameterized to match given path of output and of total separations in both models. Demand shock is assumed identical across the two models. Job-finding probability displayed only for model with selection.

Figure 6: COVID scenario - selection model (solid lines) and selection model (dashed lines). Demand shock and separation shock parameterized to match given path of output and of total separations in the selection and no-selection models. Demand shock is assumed identical across the two models.

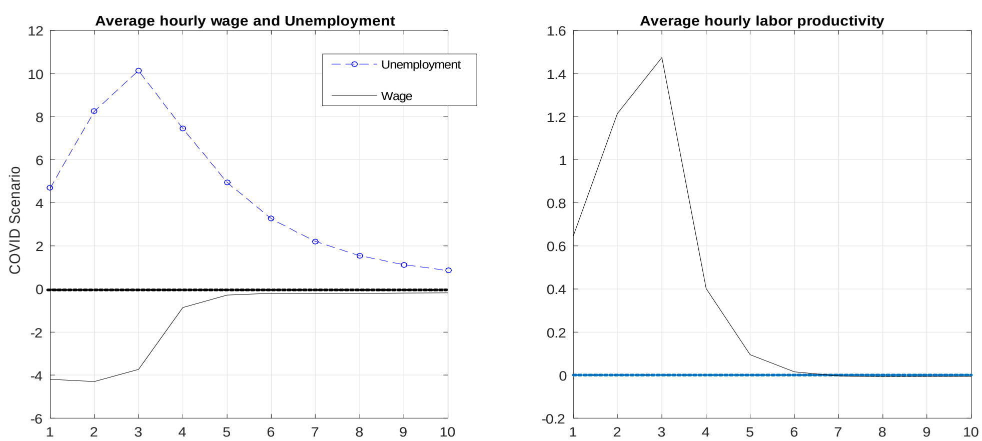

Figure 7: COVID scenario - average hourly wages, labor productivity and unemployment in the selection model.

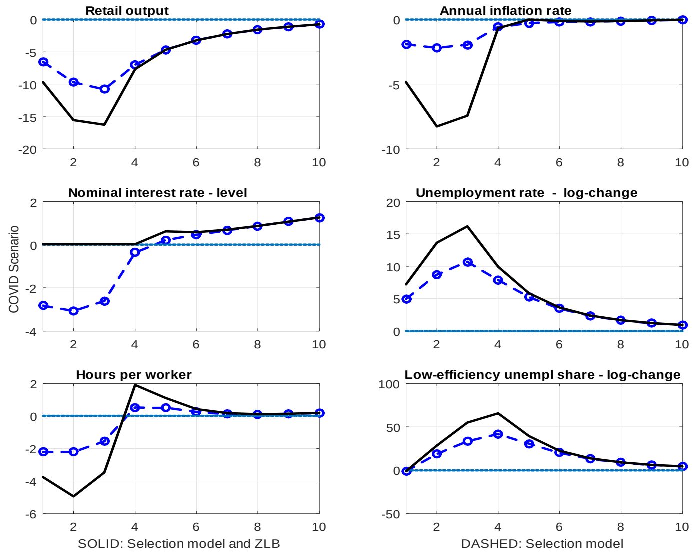

Figure 8: COVID scenario - Impact of zero lower bound on nominal interest rate in the model with selection. Dashed lines: selection model without ZLB constraint. Solid lines: selection model with the ZLB constraint. Demand shock and separation shock parameterized to match given path of output and of total separations. Zero lower bound constraint parameterized to bind for four quarters.

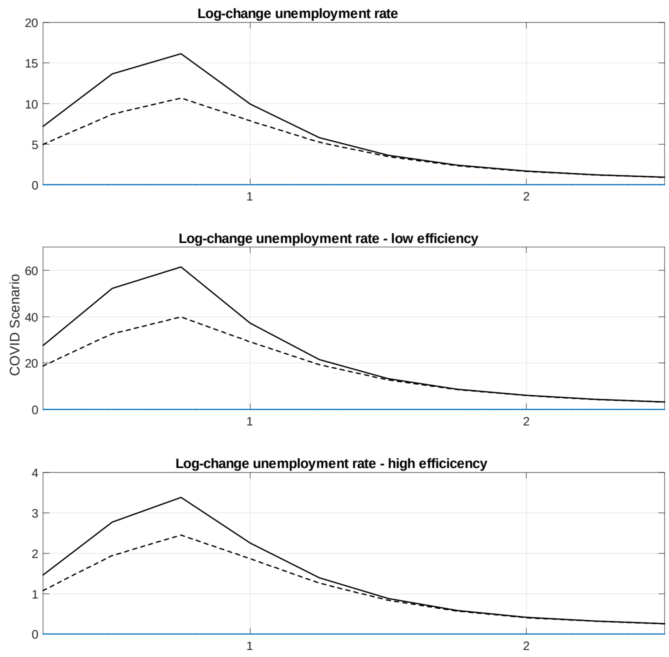

Figure 9: COVID scenario - Impact of zero lower bound on nominal interest rate in the model with selection. Dashed lines: selection model without ZLB constraint. Solid lines: selection model with the ZLB constraint. Demand shock and separation shock parameterized to match given path of output and of total separations. Zero lower bound constraint parameterized to bind for four quarters.

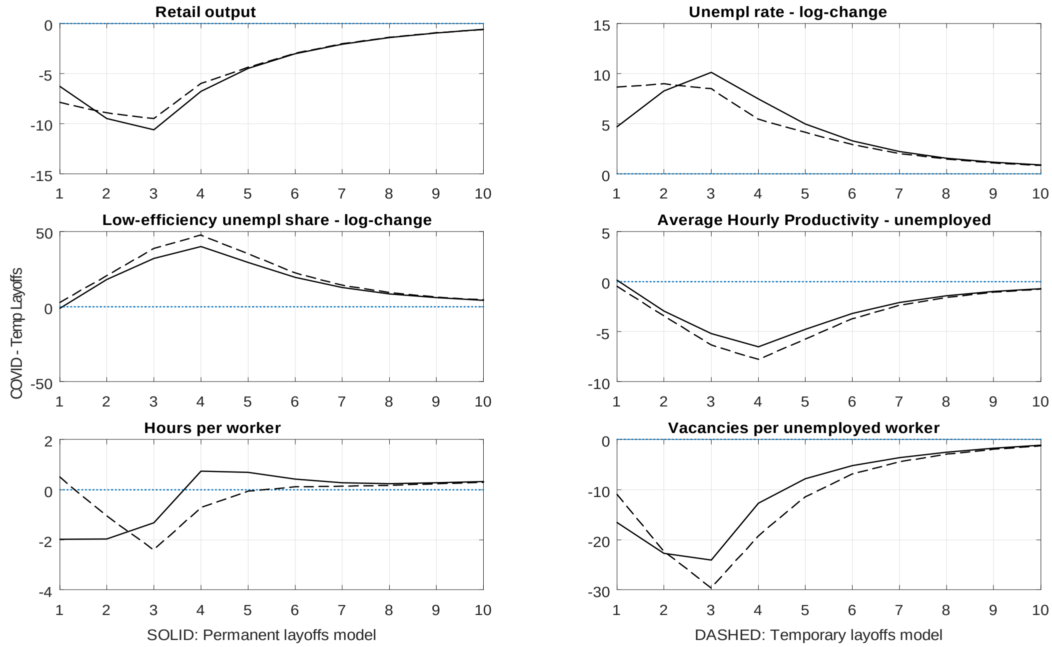

Figure 10: Covid recession with temporary layoffs. Baseline model (permanent layoffs) denoted by solid lines; model with temporary layoffs denoted by dashed lines.

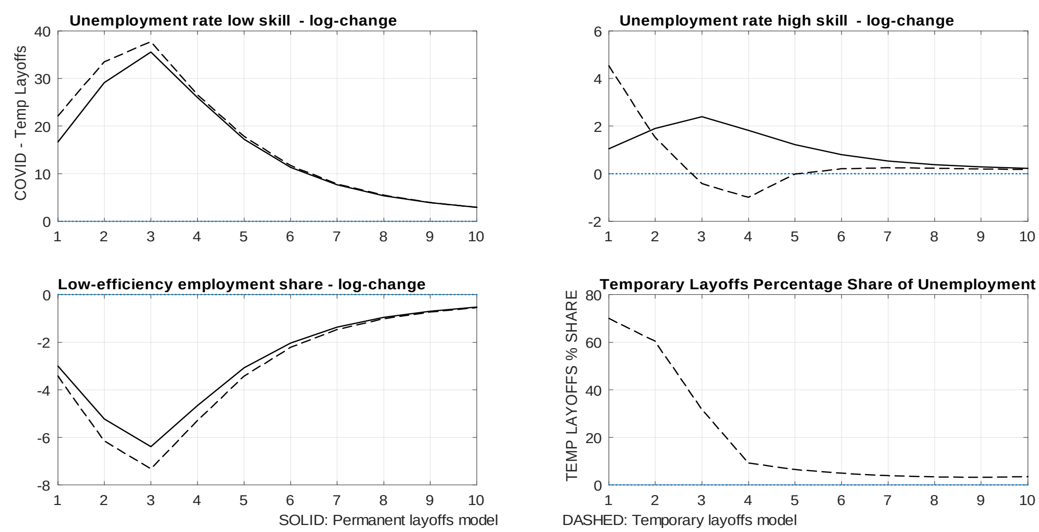

Figure 11: Covid recession with temporary layoffs. Baseline model (permanent layoffs) denoted by solid lines; model with temporary layoffs denoted by dashed lines.