# Job-to-Job Mobility and Inflation

**Renato Faccini and Leonardo Melosi**

January 19, 2023

WP 2023-03

https://doi.org/10.21033/wp-2023-03

FEDERAL RESERVE BANK *of* CHICAGO

# Job-to-Job Mobility and Inflation*

Renato Faccini
Danmarks Nationalbank
Centre for Macroeconomics (LSE)

Leonardo Melosi
FRB Chicago
CEPR

January 19 2023

**Abstract**

The low rate of inflation observed in the U.S. over the past decade is hard to reconcile with traditional measures of labor market slack. We develop a theory-based indicator of interfirm wage competition that can explain the missing inflation. Key to this result is a drop in the rate of on-the-job search, which lowers the intensity of interfirm wage competition to retain or hire workers. We estimate the on-the-job search rate from aggregate labor-market flows and show that its recent drop is corroborated by survey data. During "the great resignation", the indicator of interfirm wage competition rose, raising inflation by around 1 percentage point during most of 2021.

*Keywords*: Missing inflation, labor market slack, Phillips curve, employment-to-employment rate, micro data.

*JEL codes*: E31, E37, C32

*Correspondence to: rmmf@nationalbanken.dk and leonardo.melosi@chi.frb.org. This paper expands on work from 2020, previously circulated under the title "Bad Jobs and Low Inflation." We thank Gadi Barlevy, Robert Barsky, Marco Bassetto, Lawrence Christiano, Pablo Cuba-Borda, Martin Eichenbaum, Jason Faberman, Filippo Ferroni, Jonas Fisher, Nir Jaimovich, Michael Krause, Giuseppe Moscarini, Fabien Postel-Vinay, Sergio Rebelo, and seminar participants at the Society for Economic Dynamics, the Atlanta Fed, the Boston Fed, the Chicago Fed, the European University Institute, LSE, Northwestern University, the Bank of England, the University of Copenhagen, the University of Miami, the University of Warwick, and the EES conference in Stockholm on New Developments in the Macroeconomics of Labor Markets for their comments and suggestions. We also thank Jason Faberman for sharing the series of quit rates for the 1990s and May Tysinger and Russell Miles for providing excellent research assistance. The views in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Chicago, Danmarks Nationalbank, or any person associated with the Federal Reserve System or the European System of Central Banks.

> "Our framework for understanding inflation dynamics could be misspecified in some fundamental way, perhaps because our econometric models overlook some factor that will restrain inflation in coming years despite solid labor market conditions."
>
> Janet Yellen, Federal Reserve Chair, at the 59th Annual Meeting of the National Association for Business Economics in Cleveland, OH, on September 26, 2017

# 1 Introduction

Workhorse models used to study inflation attribute a key role to the labor market. When the labor market is tight, wage pressures and marginal costs increase, resulting in growing inflation; when the labor market is slack, wages and marginal costs fall and inflation decreases. This prediction is not borne out by the U.S. macroeconomic developments that preceded the pandemic recession, when the rate of unemployment is taken as a proxy for labor market slack, following a conventional approach dating back to Phillips (1958). As shown in Figure 1, from March 2017 through the end of 2019 the unemployment rate has stayed consistently below its average level measured over the last twelve months of the previous expansion, and by September 2019 it had reached its 50-year low at 3.5%. At the same time, core inflation according to the Price Index for Personal Consumption Expenditures (PCE) remained persistently below its long-term expectations.[1]

We first show that traditional measures of labor market slack fail to explain this missing inflation. We then introduce a more comprehensive indicator of labor-market inflationary pressures, which is related to the unemployment rate as well as the employment-to-employment (EE) flow rate. As shown in Figure 1, the EE rate has recovered sluggishly in the previous decade, remaining below its pre-Great Recession average. One factor behind this low churn rate is the decline in the propensity of employed workers to search on the job, which we show to be supported by survey data. We build a model to show that the low propensity to search on the job and the ensuing rise in low-productivity matches can explain the missing inflation in the last decade.

In the model, the productivity of jobs is match-specific and can be either high or low. All unemployed workers and a time-varying fraction of the employed search for a job. Firms have to compete to attract or retain workers who search on the job by bidding up their wage offers. As a result, these job seekers are more expensive to hire than the unemployed. A lower rate of on-the-job search reduces the incidence of wage competition between firms, leading to a

[1] In the post-war period, the U.S. economy experienced low rates of unemployment and inflation in other periods, for example in the 1960s and in the 1990s. However, these episodes occurred in connection with high labor productivity growth, which in New Keynesian (NK) models lowers real marginal costs and hence dampens inflationary pressures. What made the latest expansion particularly puzzling was that inflation remained low while labor productivity growth also slowed down (Fernald 2016).

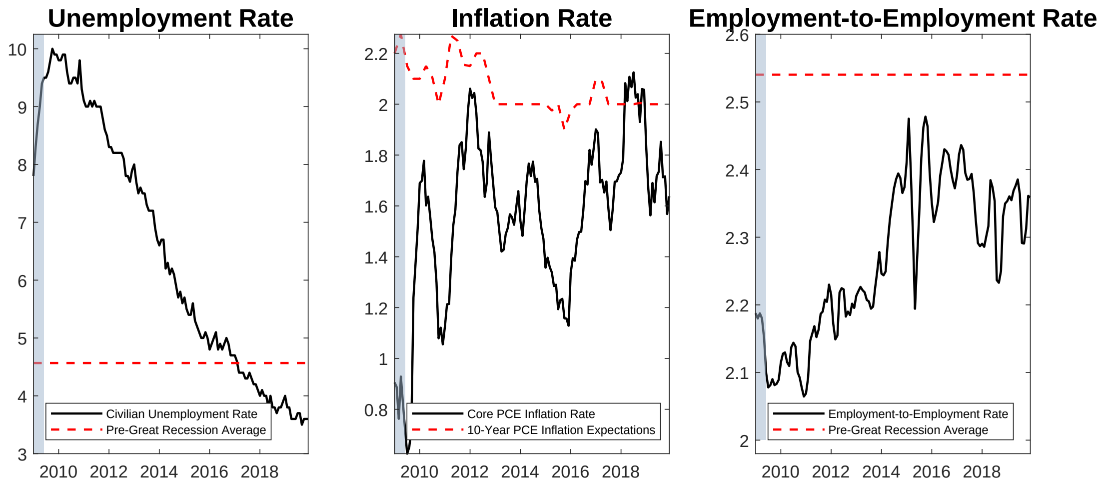


Figure 1: Labor market and inflation dynamics during the post-Great Recession recovery. The left panel: the civilian unemployment rate (solid line) and its average computed over a period spanning July 1990 through December 2007 (red dashed line). Source: Bureau of Labor Statistics (BLS). The central panel: the Price Index for Personal Consumption Expenditures (PCE) Excluding Food and Energy (black line). Average of monthly figures annualized in percentage. Source: Bureau of Economic Analysis (BEA). The red dashed line denotes the average expectations about the PCE inflation rate over the next ten years. Source: Survey of Professional Forecasters. The right panel: employment-to-employment flow rate (black line) and its average computed over a period spanning July 1990 through December 2007 (red dashed line). Three-month centered moving average. Source: Current Population Survey (CPS) extended back to January 1990 by splicing it with the quit rate measured by Davis, Faberman, and Haltiwanger (2012) and corrected by Fujita, Moscarini, and Postel-Vinay (2019). The shaded areas denote the NBER recessions.

decline in expected labor costs and lower inflationary pressures. Intuitively, if firms expect their employees to be less willing to search and quit for another job, they will also anticipate less frequent pay raise requests to match outside offers and hence less pressure on payroll costs.

We first show that the on-the-job search rate in the model is implied by the unemployment-to-employment (UE) and the EE flow rates and hence can be measured using aggregate labor market flows. A low and stagnant fraction of workers who switch jobs, combined with a tight labor market in which it becomes easier over time for the unemployed job seekers to find jobs, is interpreted by our model as a fall in the on-the-job search rate. To validate this prediction, we compute the on-the-job search rate at the micro level using a supplement to the Survey of Consumer Expectations (SCE) administered by the Federal Reserve Bank of New York. This is a new, and still largely unexplored survey, which was specifically designed to measure job search behavior, and overcome the scarcity of direct information available in alternative datasets. In the Survey, the on-the-job search rate has fallen from 2014 through 2019 in a way that is remarkably similar to our measure based on the aggregate labor market flows. While the time period covered by the survey is quite short, it contains information on the search behavior of the employed over those years that are critical for our model to explain the missing inflation.

We derive a model-consistent indicator of inflationary pressures, which can be measured using the observed series of the UE and EE rates. This indicator captures the intensity of interfirm wage competition, which is shown to depend on $(i)$ the unemployment rate, $(ii)$ the degree of cyclical labor misallocation (i.e., the incidence of low-productivity jobs), and $(iii)$ the on-the-job search rate. An increase in the rate of unemployment or a fall in the fraction of workers who are searching on the job reduces inflationary pressures in the model because they raise the firms' chances to fill their vacancies with unemployed workers, who are cheaper to hire, as they are unable to prompt wage competition between employers. The more inefficient the allocation of labor, the more likely it is for firms to meet workers employed in low-productivity (bad) matches. Because enticing a worker away from a bad match is cheaper on average than poaching a worker from a good match, labor misallocation lowers the intensity of wage competition and thereby inflationary pressures.

We then take the model to the UE and the EE flow rates observed in the data and recover the two shocks that buffet the model economy: a shock to the on-the-job search rate and a demand shock. The demand shock serves the purpose of generating the fluctuations in the UE rate observed in the data. Given the UE rate, the EE rate allows us to pin down the shocks to the on-the-job search rate, as described earlier. We use the time series of the two shocks to simulate inflation and labor costs from the model. We find that the model does not predict inflationary pressures during the most recent expansion. This result is driven by the decline in the rate of on-the-job search, which decreased wage competition both directly, and indirectly, by increasing the stock of bad matches. This prediction is consistent with the survey evidence from the SCE, which shows that in 2017, after eight years of economic recovery, about 30% of workers were not fully satisfied with how their current jobs fit their experience and skills.[2]

We then extend our analysis to the period that covers the pandemic recession and its immediate aftermaths, to see if our model can contribute to explaining the recent bout of fast growing prices, as the economic environment shifted from low to high inflation. We find that the rate of on-the-job search has increased markedly during the pandemic recession. This estimated increase in the workers' propensity to search on the job is in line with the dramatic increase in the rate at which employed workers quit their job, a phenomenon that has been dubbed "the great resignation". Our model predicts that this increase in the workers' propensity to search on the job contributed to raising inflation by around 1 percentage point during most of 2021.

Our work is related to the growing literature that has investigated why inflation has apparently become less sensitive to changes in the traditional measures of labor market slack since the early 1990s. Some scholars pointed out that the slope of the estimated price Phillips curve has flattened (Atkenson and Ohanian 2001; Stock and Watson 2007, 2008, 2019, Del Negro et al. 2020).[3] We contribute to this literature by pointing out that the most recent breakdown in the relationship between conventional measures of slack and inflation can be explained by the fall in the propensity of the employed workers to search for new jobs.

[2]In Appendix N, we show that our main results can also be obtained in a model in which the decision to search on the job is endogenous.

[3]McLeay and Tenreyro (2019) explain the identification problem posited by the estimation of Phillips curves using macrodata, showing evidence of a steeper Phillips curve in US regional data.

Our theory directly links the surprisingly low price inflation observed towards the end of the previous decade to the surprisingly low wage inflation due to weak competition among firms to hire or retain workers. An implication of this explanation is that the flattening of the price Phillips curve must be accompanied by a flattening of the wage Phillips curve. When we extend the analysis of Coibion and Gorodnichenko (2015) through the end of 2019, we document that the wage Phillips curve has flattened in the second part of the previous decade.[4] This is consistent with our evidence that the rate of on-the-job search has contributed negatively to wage and price inflation in that period.

Hazell, Herreño, Nakamura, and Steinsson (2022) find that the slope of the Phillips curve is small and has been remarkably stable over time. These scholars thereby conclude that the stability of inflation since 1990 is explained by adjustments in long-run inflation expectations becoming more firmly anchored. An implication of this conclusion is that standard labor market slack alone cannot explain the recent dynamics of inflation. Our paper is motivated by a similar consideration but our resolution to this puzzle is different in that it hinges on introducing a broader definition of labor market slack that takes into account the propensity of workers to search on the job. Ball and Mazumder (2019) show that using the short-term unemployment rate (the share of the labor force unemployed for less than 27 weeks) as a measure of slack allows an expectation-augmented price Phillips curve to explain inflation fairly well. Nevertheless, their analysis does not include the second half of the 2010s when the missing inflation puzzle arises and, unlike our study, does not provide any microfoundation for the proposed measure of slack.

Moscarini and Postel-Vinay (2019) pioneer a New Keynesian model in which cyclical labor misallocation brings about deflationary pressures. We build our model on their groundbreaking contribution. These scholars use the model to show that the degree of labor misallocation is a better predictor of inflation than the rate of unemployment. Our contribution differs from that of Moscarini and Postel-Vinay (2019) in two important respects. First, while their empirical analysis is reduced form and external to their model, our approach relies on filtering techniques to take the model to the data. Second, while Moscarini and Postel-Vinay focus exclusively on the role of cyclical labor misallocation, we emphasize the importance of the propensity to search on the job for the dynamics of wages and inflation. We show that this propensity can be measured using aggregate labor market flows and the macro estimates are validated using survey data. Crucially, allowing the on-the-job search rate to vary over time is key to explaining the missing inflation of the past decade. When the search rate is constant, the acceptance rate, which is the ratio of EE to UE flow rates, is a leading indicator for inflationary pressures. This rate is a proxy for the degree of cyclical labor misallocation and a low value of this ratio

[4] See Coibion, Gorodnichenko, and Ulate (2019) for an application of the same methodology to the analysis of the Price Phillips curve in an international context.

predicts high inflation.[5] Through the end of 2019, the acceptance rate was lower than its pre-Great Recession average in the data (see Appendix A). Our model jointly explains this low EE/UE rate, the persistent increase in bad jobs, and the low inflation in the most recent years with the decline in the incidence of on-the-job search. According to our model the EE/UE rate was low at the end of the past recovery, not because employment was efficiently allocated but because fewer workers were searching on the job.

Understanding the search behavior of the employed using disaggregated labor data is an active area of ongoing research, to which we will refer in the paper. In this paper, we stress the importance of this line of research to improve our understanding of inflation. Abraham, Haltiwanger and Rendell (2020) survey this literature and analyze the behavior of a measure of labor market tightness extended to include all effective job seekers, including employed workers searching on the job. A fall in the rate of on-the-job search, just like in any standard model, reduces labor supply and thereby increases labor market tightness, and inflationary pressures. In our model instead, a fall in the rate of on-the-job search reduces interfirm wage competition, and therefore also inflationary pressures.

The paper is organized as follows. In Section 2, we provide the motivation for our paper by laying out the missing inflation puzzle. The model from which we derive the novel indicator of wage competition is introduced in Section 3. We explain the empirical strategy and results in Section 4. Section 5 provides an empirical investigation of how the wage Phillips Curve has changed over time and interpret those changes in the light of our model. Section 6 investigates the period of the pandemic recession. In Section 7, we present our conclusions.

# 2 The Missing Inflation Puzzle

The New Keynesian model is the most popular framework to study inflation. A key building block of this framework is the New Keynesian Phillips curve, which posits that inflation $\pi_t$ hinges on the expected dynamics of current and future real marginal costs $\varphi_t$:

$$\pi_t = \kappa\varphi_t + \beta E\pi_{t+1}, \tag{1}$$

where $\kappa$ denotes the slope of the curve and $\beta$ the discount factor. In empirical applications, the real marginal cost $\varphi_t$ is proxied in a variety of ways. We consider proxies related to the following three traditional theories of the Phillips curve: ($i$) old-fashioned theories (recently revived by Galí, Smets, and Wouters 2011) that link inflation to the current and expected unemployment gap; ($ii$) the standard New Keynesian theory (derived from models with no labor frictions),

[5] The fraction of accepted offers is lower when more workers are employed in high-productivity jobs. If workers are efficiently allocated, outside offers are declined and matched by the current employer, raising production costs and inflation.

which suggests that the labor share alone is the key determinant of the inflation rate (Galí and Gertler 2000); (*iii*) a variant of the standard New Keynesian theory, based on models that account for search and matching frictions, which explains inflation using current and expected measures of the labor share as well as UE flow rates (Krause, Lopez-Salido, and Lubik, 2008).[6]

While there are more sophisticated versions of the New Keynesian Phillips curve, which, for instance, feature price indexation, we focus here on the simpler version of this curve to facilitate comparability with the model presented in the next section. We discuss the extension to the case of price indexation in Appendix C and show that it does not affect our main conclusions.

By solving equation (1) forward, we can express expected inflation as the sum of the current and future expected real marginal costs. We estimate a Vector Autoregression (VAR) model to forecast the future stream of the three aforementioned measures of real marginal costs. The forecasts of real marginal costs are launched from every quarter during the post-Great Recession recovery and are then plugged into the Phillips curve, which returns the predicted inflation rate by each of the three theories of marginal costs in every quarter of the recovery. To conduct this exercise, we set the discount factor $\beta$ to 0.99 (data are quarterly) and a slope of the Phillips curve $\kappa$ equal to 0.005, so as to fit inflation at the beginning of the post-Great Recession recovery (2009–2011). The resulting Phillips curve is fairly flat and in line with estimates obtained by Del Negro et al. (2020) for the U.S economy over the post-1990 period, using standard measures of slack. While the slope of the Phillips curve affects the magnitude of inflation predicted by the three measures of slacks, it does not affect the point in time when inflation rises above its long-run level, which is what we are interested in (see Appendix C). A flatter Phillips curve in and of itself does not explain why inflation has remained subdued for an entire decade.

To estimate the VAR model, we use the following observable variables: the labor share, the job finding rate, real wages, the civilian unemployment rate, real gross domestic product (GDP), real consumption, real investment, inflation according to the Consumer Price Index (CPI), and the federal funds rate (FFR).[7] We detrend the observable variables by using their 8-year past moving average trend. The only exception is when we construct the unemployment gap, for which we use the short-term NAIRU estimates.[8] We rely on the NAIRU estimates to construct the unemployment gap as this practice is very popular in those studies whose object is to estimate the Phillips curve. The sample period for estimation is from 1958Q4 through 2017Q4.

Figure 2 shows that all the three traditional theories of marginal costs predict that inflation should have been above its long-run level (positive inflation gap) by the end of 2012. None of these theories is able to account for why inflation stayed so low for so many years after the

[6]To make the paper self-contained, we summarize how this third series of marginal costs is constructed in Appendix B. We refer the interested reader to Krause, Lopez-Salido, and Lubik (2008) for more details.

[7]Details on how these series are constructed are in Appendix D.

[8]Using the long-term NAIRU would not change our main conclusions.

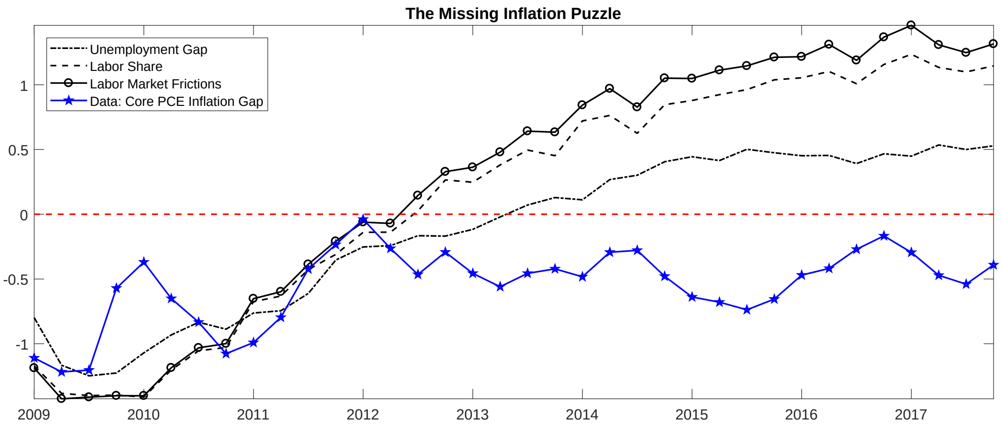


Figure 2: Core PCE inflation gap from 2009Q1 through 2017Q3 and inflation dynamics predicted using three traditional theories of inflation.

Great Recession because all three proxies for marginal costs improved quickly in the first years of the economic recovery. Consequently, the VAR model's forecasts of future marginal costs go up at a relatively early stage of the recovery, which leads the three New Keynesian Phillips curves to predict inflation above its long-run level. As shown in Appendix E, a state-of-the-art structural model, such as the model studied in Smets and Wouters (2007), also fails to explain the missing inflation.

It is worth noting that this VAR approach is general and agnostic because we do not impose a requirement that expectations about future labor costs are formed according to the Phillips curve.[9] Imposing such a restriction on the VAR structure may lead to misspecification that would most likely worsen the quality of the forecasts of real marginal costs. That said, our main conclusions are not affected by imposing this restriction, and our approach is more appealing in that the unrestricted VAR model is a reduced-form, theory-free representation for the data that is less prone to misspecification than structural theory-based models.

# 3 A General Equilibrium Model with On-the-Job Search

The failure of the traditional measures of labor market slack to explain inflation in the past decade motivates the need of an alternative indicator of inflationary pressures. We build one using a New Keynesian model in which a time-varying fraction of workers search on the job and firms have to compete to attract or retain these workers by bidding up their wage offers.[10]

[9] Instead, the forecasts are provided by the estimated Bayesian VAR (BVAR) model. BVAR models are known to provide good macroeconomic forecasts if the direction of prior shrinkage is carefully determined. We do so by following the forecasting literature. Specifically, we use the unit-root prior introduced by Sims and Zha (1998) and choose the prior hyperparameters, which determine the direction of the Bayesian shrinkage, so as to maximize the marginal likelihood. We also check that our VAR forecasts are accurate in sample.

[10] Key empirical studies that explicitly allow for search and matching frictions in New Keynesian models include Gertler, Sala, and Trigari (2008), Krause, Lopez-Salido, and Lubik (2008), Ravenna and Walsh (2008)

## 3.1 The Economy

The economy is populated by a representative, infinitely lived household, whose members are either unemployed or employed. All members of the household pool their income at the end of each period and thereby consume the same amount. The labor market is frictional and workers search for jobs whether they are unemployed or employed. While all unemployed workers are also job seekers, any employed worker can search in a given period with probability $s_t$, which is assumed to follow an exogenous first-order autoregressive AR(1) process with Gaussian shocks. Time variation in $s_t$ is meant to capture all those cyclical factors that are responsible for changes in the average rate of on-the-job search in the data, including compositional changes in the propensity to search in the pool of employed workers. Households trade one-period-government bonds $B_t$.

We distinguish two types of firms: labor-service producers and price setters. The service sector comprises an endogenous measure of worker–firm pairs that match in a frictional labor market and produce a homogeneous nonstorable good. Productivity $y \in \{y_g, y_b\}$ is match-specific and can be either good or bad, with $y_g > y_b > 0$. We let $\xi_g$ denote the probability that upon matching the productivity draw is good and $\xi_b = 1 - \xi_g$ the probability that the draw is bad. The output of the match is sold to price-setting firms in a competitive market at the relative price $\varphi_t$, and transformed into a differentiated product. Specifically, one unit of the service is transformed by firm $i$ into one unit of a differentiated good $y_t(i)$. These firms set the price of their goods subject to Calvo price rigidities. Households consume a bundle $C_t$ of such varieties in order to minimize expenditure. This bundle is the numeraire for this economy and its price is denoted by $P_t$. The monetary authority sets the nominal interest rate of the one-period government bond following a Taylor rule subject to a nonnegativity constraint. The fiscal authority levies lump-sum taxes $T_t$ to finance maturing government bonds.

## 3.2 The Labor Market

The labor market is frictional and governed by a meeting function that brings together vacancies and job seekers. The pool of workers looking for jobs at each period of time $t$ is given by the measure of workers who are unemployed at the beginning of a period, $u_{0,t}$ plus a fraction $s_t$ of the workers who are employed, $1 - u_{0,t}$. So following Moscarini and Postel-Vinay (2019), we interpret $s_t(1 - u_{0,t})$ as the measure of employed workers who search on the job. Denoting the aggregate mass of vacancies by $v_t$, we can define labor market tightness as

and Christiano, Eichenbaum, and Trabandt (2016). We deviate from these studies by considering the role of on-the-job search and by focusing on inflation.

Gertler, Huckfeldt, and Trigari (2019) develop a model where productivity is match specific, and workers climb the ladder by searching on the job. Their paper abstracts from nominal rigidities and focuses on the wage cyclicality of the newly hired workers.

$$\theta_t = \frac{v_t}{u_{0,t} + s_t (1 - u_{0,t})}. \tag{2}$$

We assume that the meeting function is homothetic, which implies that the rate at which searching workers find a vacancy, $\phi(\theta) \in [0,1]$, and the rate at which vacancies draw job seekers, $\phi(\theta)/\theta \in [0,1]$, depend exclusively on $\theta$ and are such that $d\phi(\theta)/d\theta > 0$ and $d[\phi(\theta)/\theta]/d\theta < 0$.

Because of frictions in the labor market, wages deviate from the competitive solution. It is assumed that wage bargaining follows the sequential auction protocol of Postel-Vinay and Robin (2002). Namely, the outcome of the bargaining is a wage contract, i.e., a sequence of state-contingent wages, which promises to pay a given utility payoff in expected present value terms (accounting also for expected utility from future spells of unemployment and wages paid by future employers). The commitment of the worker-firm pair to the contract is limited, in the sense that either party can unilaterally break up the match if either the present value of firm profits becomes negative or the present value utility from being employed falls below the value of being unemployed. The contract can be renegotiated only by mutual consent: if an employed worker meets a vacancy, the current and the prospective employer observe first the productivity associated with both matches, and then engage in Bertrand competition over contracts. The worker chooses the contract that delivers the largest value.

The within-period timing of actions is as follows: all the unemployed workers and a fraction $s_t$ of the employed search for a job at the beginning of the period. Next, some workers move out of the unemployment pool, while successful on-the-job seekers have their wage renegotiated and possibly move up the ladder. Then production takes place and wages are paid. This timing implies that workers who are unemployed at the beginning of the period can produce at the end of the same period if they find a job. And similarly, workers who are employed at the beginning of the period may be producing in a different job at the end of the same period if they switch employers. Finally, a fraction $\delta$ of the existing matches is destroyed.

These assumptions imply the following dynamics for the aggregate state of unemployment. Denote the stock of end-of-period employed workers as

$$n_t = 1 - u_t. \tag{3}$$

Aggregate unemployment at the beginning of a period is given by

$$u_{0,t} = u_{t-1} + \delta n_{t-1}, \tag{4}$$

while aggregate unemployment at the end of a period is given by

$$u_t = u_{0,t} [1 - \phi(\theta_t)]. \tag{5}$$

### 3.3 Households

Households solve two problems. First, they decide how to optimally allocate their consumption of the aggregate good over time. Second, they solve an intratemporal problem to optimally choose the composition of the aggregate good in terms of differentiated goods sold by the price setters.

**The intertemporal maximization problem** The representative household enjoys utility from the consumption basket $C_t$ and from the fraction of its members who are not working and are therefore free to enjoy leisure. The parameter $b$ controls the marginal utility of leisure. We assume that the utility function is logarithmic in consumption and let $\mu_t$ denote the preference shock to consumption, which is assumed to follow a Gaussian AR(1) stochastic process in logs. The resources available to consume at a given point in time $t$ include government bond holdings $B_t$; profits from the price setters, which sell differentiated goods, $D_t^P$; profits from the service firms $D_t^S$; wages from the workers who are employed; and transfers from the government $T_t$.

We assume that all unemployed workers look for jobs, and restrict our attention to equilibria where the value of being employed for any worker is no less than the value of being unemployed. In this setup, the measure of workers who are employed is not a choice variable of the household, but is driven by aggregate labor market conditions through the job finding probability $\phi(\theta_t)$. Let $e_t(j) \in \{0, 1\}$ be an indicator function which takes the value of one if a worker $j$ is employed after worker reallocation takes place, but before the current-period exogenous separation occurs with probability, $\delta$, and zero otherwise. The intertemporal maximization problem is

$$\max_{\{C_t, B_{t+1}\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[\mu_t \ln C_t + b \int_0^1 (1 - e_t(j))\, dj\right],$$

subject to the budget constraint,

$$P_t C_t + \frac{B_{t+1}}{1 + R_t} \leq B_t + \int_0^1 e_t(j)\, w_t(j) + D_t^P + D_t^S + T_t,$$

and the stochastic processes for the employment status,

$$\begin{aligned} prob\{e_{t+1}(j) = 1 \mid e_t(j)\} &= e_t(j)\left[(1-\delta) + \delta\phi(\theta_{t+1})\right] + \left[1 - e_t(j)\right]\phi(\theta_{t+1}) \\ prob\{e_{t+1}(j) = 0 \mid e_t(j)\} &= 1 - prob\{e_{t+1}(j) = 1 \mid e_t(j)\}, \end{aligned} \tag{6}$$

and for equilibrium wages $w_t(j)$.[11]

[11] The evolution of individual wages must satisfy the wage contract negotiated by the worker-firm pair. In these negotiations, workers and firms agree on a present discounted value of the future stream of utility. Because there are many streams of wages that can deliver the promised present discounted value of utility, the distribution of individual wages is indeterminate. This indeterminacy is inconsequential for aggregate equilibrium outcomes.

Equation (6) implies that a worker $j$ who is registered as unemployed at the production stage of period $t$, i.e., $e_t(j) = 0$, will only have a chance to look for jobs at the beginning of next period, and get one with probability $\phi(\theta_{t+1})$. Moreover, a worker employed at time $t$, i.e., $e_t(j) = 1$, will also be in employment at $t+1$ if she does not separate from the current job between periods at the exogenous rate $\delta$, or if she separates but manages to find a new job with probability $\phi(\theta_{t+1})$ in the next period.

**The intratemporal minimization problem conditions** Households minimize total expenditure on all differentiated goods,

$$\min_{q_t(i), i\in[0,1]} \int_0^1 p_t(i)\, q_t(i)\, di, \tag{7}$$

subject to the general Kimball (1995) aggregator assumed in Smets and Wouters (2007):

$$\int_0^1 G\left(q_t(i)/Q_t\right) di = 1. \tag{8}$$

The reason why we choose this particular aggregator will be explained in Section 4.1, where we discuss how we calibrate the key parameter of this aggregation technology. As in Dotsey and King (2005), Levin, Lopez-Salido, and Yun (2007), and Lindé and Trabandt (2018), we assume the following strictly concave and increasing function for $G(q_t(i)/Q_t)$:

$$G\left(q_t(i)/Q_t\right) = \frac{\omega^k}{1+\varkappa}\left[(1+\varkappa)\frac{q_t(i)}{Q_t} - \varkappa\right]^{\frac{1}{\omega^k}} + 1 - \frac{\omega^k}{1+\varkappa}, \tag{9}$$

where $\omega^k = \frac{\chi(1+\varkappa)}{1+\varkappa\chi}$, $\varkappa \leq 0$ is a parameter that governs the degree of curvature of the demand curve for the differentiated goods and $\chi$ captures the gross markup.

The solution of this expenditure minimization problem is the demand function for the differentiated good $(i)$:

$$\frac{q_t(i)}{Q_t} = \frac{1}{1+\varkappa}\left(\frac{P_t(i)}{P_t\Xi_t}\right)^{\iota} + \frac{\varkappa}{1+\varkappa}, \tag{10}$$

where $\varkappa \leq 0$ is a parameter, $\iota = \frac{\chi(1+\varkappa)}{1-\chi}$, $\Xi_t$ is the Lagrange multiplier associated with the constraint (8), and the aggregate price index (i.e., the price of the numeraire) satisfies $1 = \int_0^1 \left(\frac{p_{t,i}}{P_t\Xi_t}\right)^{\frac{\iota}{\omega^k}} di$.

Moreover, as we will clarify later, the real marginal cost, which is the price of the labor service and hence a measure of the average cost of labor, is determined, even though the underyling wage distribution is not.

### 3.4 Price Setters

Price setters buy the (homogeneous) output produced by the service firms in a competitive market at the relative price $\varphi_t$, turn it into a differentiated good, and sell it to the households in a monopolistic competitive market. They can re-optimize their price $P_t(i)$ with a period probability $1-\zeta$. If they cannot reoptimize, they adjust their price at the steady-state inflation rate $\Pi$. Therefore, the problem of the price setting firm is expressed as follows:

$$\max_{P_{t+s}(i)} E_t \sum_{s=0}^{\infty} \beta^{t+s} \zeta^s \frac{\lambda_{t+s}}{\lambda_t} \left( P_t(i) \Pi^s - P_{t+s} \varphi_{t+s} \right) q_{t+s}(i), \tag{11}$$

subject to the demand function (10). Log-linearization and standard manipulations of the resulting price-setting equation lead to the purely forward-looking New Keynesian Phillips curve, which was shown in equation (1).

As standard in New Keynesian models, the Calvo lottery makes this price-setting problem dynamic; i.e., price setters that are allowed to re-optimize their price at time $t$ find it optimal to forecast the future stream of real marginal costs $\{\varphi_\tau\}_{\tau=t}^{\infty}$. This is because price setters anticipate that they may not be able to re-optimize their price in the next periods. In our model, the price setters' real marginal costs $\varphi_t$ coincide with the relative price of the labor service, and hence, the optimizing price setters care about the determinants of that price, which are the focus of the next section.

### 3.5 Service Sector Firms: Free-Entry Condition

In this section, we introduce the free-entry condition to the labor service sector and discuss the pivotal role it plays in determining the dynamics of price setters' marginal costs and inflation. This condition implies that entrant firms will make zero profits in expectations; i.e., expected costs are equal to the expected surplus after the match is formed. We discuss first the expected costs, then the expected surplus.

Service firms have to pay an advertising cost $c$ per period. In addition, to form a match and produce, they also have to pay a sunk fixed cost of hiring $c^f$. The expected cost of creating a job equals $c^f + \frac{c}{\varpi_t}$, where $\varpi_t$ is the vacancy filling rate and $\varpi^{-1}$ measures the expected number of periods that are required to meet a worker.

The expected return from a match depends on whether the worker matched is employed or unemployed. Following Postel-Vinay and Robin (2002) and Moscarini and Postel-Vinay (2018, 2019), it is assumed that unemployed workers have no bargaining power, so the firm will appropriate the entire surplus of the match, which will in turn depend on its quality. If the firm meets an employed worker instead, the firm engages in Bertrand competition with the incumbent firm in an attempt to poach the worker away from the current match. An important implication of assuming Bertrand competition is that an increase in wages is not necessarily backed by a rise in workers' productivity. This can happen, for instance, when a worker renegotiates upward the value of a contract, as their employer agrees to match the offer of a poaching firm. This temporary decoupling between wages and the worker's productivity is key for the job ladder to have meaningful implications for inflation.

While the assumption that unemployed workers have no bargaining power is not necessary to solve the model and to study its empirical properties, it provides tractability, allowing for an analytical characterization of the expected surpluses that appear in the free-entry condition. Such an analytical characterization turns out to be very useful in providing intuition about the link between the labor market and inflation in the model, which will be the focus of Section 3.6.

Importantly, the Postel-Vinay and Robin's bargaining protocol breaks the direct link between labor market tightness and wages, allowing us to isolate the effects of searching on the job on firms' wage competition. Specifically, a drop in the on-the-job search rate leads to a fall in the share of job seekers that are employed, thereby reducing interfirm wage competition and the cost of the labor service. This is the first paper that focuses on this channel to explain inflation. It is worth noting that the assumption of Bertrand competition between employers reverses the implications of on-the-job search for inflation, relative to models with Nash bargaining. Indeed, in the latter case a fall in the rate of on-the-job search reduces the number of job seekers, increasing labor market tightness and wages. For instance, this channel is emphasized in Abraham, Haltiwanger, and Rendell (2020) and in most of the reduced-form labor literature reviewed in that paper. Our model-based empirical analysis is therefore constructed to derive a simple measure of inflationary pressure that isolates the role of interfirm wage competition and investigate how far it can go in explaining the dynamics of inflation.

To illustrate how Bertrand competition works in our model, let $y$ and $y'$ denote the match quality with the incumbent and the poaching firm, respectively. We distinguish three possible contingencies.

1. $y = y_g$ and $y' = y_b$. In this case the poaching firm is a worse match for the worker. Bertrand competition implies that the incumbent firm will retain the worker and poaching is not successful. If the worker was hired from the state of unemployment, she appropriates the surplus $S_t(y_b)$ because Bertrand competition forces the incumbent to pay the worker the highest value the poaching firm is willing to pay her. If the worker was not hired from a state of unemployment, there is no change in the value of her contract.

2. $y = y'$ for $y \in \{y_b, y_g\}$. Match quality is the same for the two firms, and the worker will be indifferent between switching jobs or staying. We assume that switching takes place with probability $\nu$ (a nonzero value for this parameter is required to match the high churning

rate in the U.S. labor market when calibrating the model's steady-state parameters)[12]. In either case, the firm that ends up with the worker relinquishes all the surplus $S_t(y)$.

3. $y = y_b$ and $y' = y_g$. Match quality is lower with the incumbent firm, so the worker is poached. Bertrand competition implies that the worker is given the highest surplus the incumbent firm is willing to pay her, i.e., $S_t(y_b)$. The poaching firm's surplus is therefore the residual value of the match: $S_t(y_g) - S_t(y_b)$.

To sum up, entrant labor service firms can get a nonzero surplus from meeting an employed worker only if the worker is in a bad match and the firm is a good match for the worker. As a result, the free-entry condition can be written as follows:

$$\begin{aligned} c^f + \frac{c}{\varpi_t} = & \frac{u_{0,t}}{u_{0,t} + s_t(1 - u_{0,t})} \left\{ \xi_b S_t(y_b) + \xi_g S_t(y_g) \right\} \\ & + \frac{s_t(1 - u_{0,t})}{u_{0,t} + s_t(1 - u_{0,t})} \left\{ \xi_g \frac{l^0_{b,t}}{1 - u_{0,t}} \left[ S_t(y_g) - S_t(y_b) \right] \right\}, \end{aligned} \tag{12}$$

where $l^0_{b,t}$ denotes the measure of workers who, at the beginning of period $t$, are employed in low-quality matches ($l^0_{b,t} + l^0_{g,t} + u_{0,t} = 1$). The term $s_t(1 - u_{0,t})$ denotes the measure of employed workers searching on the job at the beginning of period $t$, and $u_{0,t} + s_t(1 - u_{0,t})$ is the measure of all job seekers at the beginning of period $t$.

The left-hand side is the expected costs of posting a vacancy, which has been discussed above. The expected return from forming a match, on the right-hand side, depends on the employment status, on the quality of the meeting, and, in the case the firm meets an employed worker, also on the quality of the existing match. Three contingencies will give a nonzero surplus to the firm and will hence appear in the right-hand side of the free-entry equation (12). The expected return on the right-hand side is the average of the surplus accrued in these three contingencies weighted by their respective probabilities.

The first contingency is when the entrant firm meets an unemployed job seeker, with probability $u_{0,t}/[u_{0,t} + s_t(1 - u_{0,t})]$, and the job seeker is a bad match for the firm, with probability $\xi_b$. In this case the meeting gives the firm the surplus $S_t(y_b)$. The second contingency is when the entrant firm meets an unemployed job seeker who turns out to be a good match, with probability $\xi_g$, providing the firm with the surplus $S_t(y_g)$. These two expected returns appear in the first term on the right-hand side of the free-entry equation (12). The third contingency, i.e., the second term in the right-hand side of the free-entry equation (12), occurs when the firm

[12] Allowing for $\nu > 0$ is consistent with evidence in Table 4 of Haltiwanger, Hyatt, Kahn, and McEntarfer (2018), who show that a large fraction of EE moves are not associated with movements up the ladder, but rather lateral moves to jobs of similar productivity. The same table also shows that these lateral moves are much more frequent than moves down the ladder, which do not feature in our model. It should be noted that the latter type of movements may be inflationary or deflationary depending on the extent to which productivity falls in the new match, relative to the negotiated wage.

meets an employed worker, with probability $s_t(1-u_{0,t})/[u_{0,t}+s_t(1-u_{0,t})]$, and the following two conditions are met: ($i$) the worker is a good match for the entrant firm, which occurs with probability $\xi_g$, and ($ii$) the worker is currently in a bad match, which happens with probability $\xi_g l^0_{b,t}/(1-u_{0,t})$.[13]

Moscarini and Postel-Vinay (2019) show that the surplus function can be written as follows

$$S_t(y) = y\mathcal{W}_t - \frac{b\lambda_t^{-1}}{1-\beta(1-\delta)}, \tag{13}$$

where $\lambda_t$ is the Lagrange multiplier associated with the household's budget constraint and

$$\mathcal{W}_t = \varphi_t + \beta(1-\delta)E_t\frac{\lambda_{t+1}}{\lambda_t}\mathcal{W}_{t+1}. \tag{14}$$

See Appendix F for details on the derivations in the context of our model.

### 3.6 Interfirm Wage Competition and Inflation

In this section, we discuss the determinants of inflation in the model. To this end, it is convenient to rewrite the free-entry condition (12) after replacing the surplus functions (13), which yields:

$$\begin{aligned} c^f + \frac{c}{\varpi_t} \quad &= \quad \frac{u_{0,t}}{u_{0,t}+s_t(1-u_{0,t})}\left[\mathcal{W}_t\left(\xi_b y_b + \xi_g y_g\right) - \frac{b\lambda_t^{-1}}{1-\beta(1-\delta)}\right] \\ &\quad + \frac{s_t(1-u_{0,t})}{u_{0,t}+s_t(1-u_{0,t})}\xi_g\frac{l^0_{b,t}}{1-u_{0,t}}\left[\mathcal{W}_t(y_g-y_b)\right]. \end{aligned} \tag{15}$$

This equation implicitly links the expected present discounted value of the entire stream of current and future real marginal costs, $\mathcal{W}_t$, with a set of labor market variables which include the unemployment rate ($u_{0,t}$), the measure of misallocated workers ($l^0_{b,t}$), the on-the-job-search rate ($s_t$), and the vacancy-filling rate ($\omega_t$). We notice that both $\mathcal{W}_t$ and inflation ($\pi_t$) are discounted values of current and future real marginal costs of price setters (i.e., the relative price of the homoegenous good sold by the labor service firms).

To understand how the free entry condition works and the role it plays in affecting inflation dynamics, it is useful to define $\Sigma_t$ to be the probability that, conditional on a contact, firms entering the labor service sector are not engaged in a wage competition that leads them to relinquish the entire surplus to the worker. This probability is defined as follows:

$$\Sigma_t \equiv \frac{u_{0,t}}{u_{0,t}+s_t(1-u_{0,t})} + \frac{s_t(1-u_{0,t})}{u_{0,t}+s_t(1-u_{0,t})}\xi_g\frac{l^0_{b,t}}{1-u_{0,t}}, \tag{16}$$

[13] Note that $l^0_{b,t}$ denotes the share of workers that are employed in a bad match at the beginning of the period. We rescale this share by the fraction of employed workers at the beginning of the period $(1-u_{0,t})$ so as to obtain the conditional probability of meeting a bad match.

where the first term on the right-hand side is the probability of meeting an unemployed worker and the second term is the probability of meeting a worker employed in a bad match, who is searching on the job, and turns out to be a good match for the poaching firm. Notice that these two terms correspond to the expressions that premultiply the terms in the square brackets on the right-hand side of the free entry condition—equation (15). As such, the probability $1 - \Sigma_t$ measures the complement probability that a firm relinquishes the entire match surplus as a result of interfirm wage competition. In the rest of the paper, we use $1 - \Sigma_t$ as an indicator of wage competition.

Next, we provide a heuristic explanation for why a high intensity of wage competition $(1-\Sigma_t)$ leads to upward pressures on price setters' marginal costs and hence on inflation.

Let us assume that, everything else equal, either the unemployment rate $u_t^0$ increases, the fraction of workers in a bad match $l_{b,t}^0$ rises, or employed workers search less frequently ($s_t$ decreases). The direct effect of these changes on the free-entry condition is to lower the intensity of interfirm wage competition, $1 - \Sigma_t$, and, consequently, to increase firms' expected profits—i.e., the right-hand side of the free entry condition. Intuitively, any of these changes implies a fall in the expected cost of labor. An increase in $u_0$ or a fall in $s_t$ imply a higher probability that vacancies meet with workers who are unemployed, and therefore cheaper to hire. A higher $l_{b,0}$ imply a higher probability of meeting with a worker employed in a bad match, conditional on meeting with an employed. In turn mismatched workers are cheaper to hire, since their employer cannot possibly offer a contract with a value higher than $S(y_b)$.

How is the zero profit condition—equation (15)—restored following a decline in the intensity of wage competition, $1 - \Sigma_t$ and the ensuing increase in expected profits? On the one hand, the increase in vacancies leads to a decrease in the vacancy filling rate, $\varpi_t$, and to an increase in the expected cost of entry (i.e., the left-hand side of equation (15). On the other hand, the expected discounted stream of the relative prices of the labor service $\mathcal{W}_t$ falls, lowering surpluses $S_t(y)$, as implied by equation (13), and further dissuading firms from posting new vacancies until the zero-profit condition is restored.[14]

By lowering the discounted relative price of the labor service, this second rebalancing channel causes price setters' current and expected marginal costs to fall, lowering inflation. Consequently, this second channel traces an important link between inflation and the key variables driving the intensity of interfirm wage competition $(1 - \Sigma_t)$ shown in equation (16)—i.e., the unemployment rate, $u_{0,t}$, the share of low-productivity jobs, $l_{b,t}^0$, and the on-the-jobs search rate, $s_t$.

In this section, we heuristically explained the mechanics of how labor market state variables affect the intensity of inter-firm wage competition and inflation in the model. In Appendix G,

[14] With flexible prices, real marginal costs are constant and the equilibrium of the free-entry condition is restored only through a change in the vacancy filling rate.

we show analytically that—under certain parametric restrictions—inflation can be explained solely as a function of the three key drivers of the intensity of wage competition, $1 - \Sigma_t$. In Section 4.4, we will show that these parametric restrictions are approximately satisfied in the calibrated model.

## 3.7 The Dynamic Distribution of Match Types

The laws of motion for bad and good matches are

$$l_{b,t} = \left[1 - s_t \phi\left(\theta_t\right) \xi_g\right] l^0_{b,t} + \phi\left(\theta_t\right) \xi_b u_{0,t}, \tag{17}$$

$$l_{g,t} = l^0_{g,t} + s_t \phi\left(\theta_t\right) \xi_g l^0_{b,t} + \phi\left(\theta_t\right) \xi_g u_{0,t}. \tag{18}$$

In the above equations, we let $l_{b,t}$ and $l_{g,t}$ denote the end-of-period measure of bad and good matches, respectively. We let $l^0_{b,t}$ and $l^0_{g,t}$ denote beginning-of-period values. In turn, $l_{b,t}$ is equal to the sum of the bad matches at the beginning of a period that did not move up the ladder by finding a high-quality match within the period, $\left[1 - s_t \phi\left(\theta_t\right) \xi_g\right] l^0_{b,t}$, plus the new hires from the unemployment pool who turned out to draw a low-quality match, $\phi\left(\theta_t\right) \xi_b u_{0,t}$. Indeed, job-to-job flows from bad- to good-quality matches are given by the fraction of badly matched employed workers, $l^0_{b,t}$, who search on the job with exogenous probability $s_t$, meet a vacancy with probability $\phi\left(\theta_t\right)$, and draw a good match with probability $\xi_g$.

The end-of-period measure of good matches is instead given by the beginning-of-period measure of good matches $l^0_{g,t}$, plus the job-to-job inflows from bad matches $s_t \phi\left(\theta_t\right) \xi_g l^0_{b,t}$, and the unemployed hired in a good job, $\phi\left(\theta_t\right) \xi_g u_{0,t}$. Using the identity $l^0_{i,t+1}(y) = (1 - \delta)\, l_{i,t}(y)$ for $i = \{b, g\}$, to replace for $l_{i,t}(y)$ on the left-hand side of equations (17) and (18) delivers laws of motion for bad and good jobs at their beginning-of-period values.

## 3.8 Policymakers and Market Clearing

The fiscal authority levies lump-sum taxes to repay its maturing bonds in every period. The monetary authority follows a Taylor rule when the nominal interest rate $R_t$ is not constrained by the zero lower bound:

$$\frac{R_t}{R^*} = \max\left\{\frac{1}{R^*}, \left(\frac{R_{t-1}}{R^*}\right)^{\rho_r}\left[\left(\frac{\Pi_t}{\Pi^*}\right)^{\phi_\pi}\left(\frac{Q_t}{Q^*}\right)^{\phi_y}\right]^{1-\rho_r}\right\}, \tag{19}$$

where $\frac{1}{R^*}$ represents the lower bound of the nominal interest rate, $\rho_r \in [0, 1)$ captures the degree of interest rate smoothing, and the parameters $\phi_\pi > 1$ and $\phi_y > 0$ capture how strongly the central bank responds to inflation (in deviation from the target $\Pi^*$) and output (in deviation from its potential level $Q^*$). We do not include monetary shocks in equation (19) because these shocks cannot be separately identified by preference shocks in our empirical analysis.

Market clearing in the market of price-setting firms implies that the quantity sold summing over all producers $i$ must be equal to the production in the service sector:

$$y_g l_{g,t} + y_b l_{b,t} = \int_0^1 q_t(i)\, di.$$

In turn, aggregate output from price setters must equal aggregate demand from the households:

$$\int_0^1 q_t(i)\, di = Q_t \int_0^1 \left( \frac{1}{1+\varkappa} \left( \frac{P_t(i)}{P_t \Xi_t} \right)^{\iota} + \frac{\varkappa}{1+\varkappa} \right) di,$$

where we have made use of the demand function in equation (10). Substituting the profits of all firms into the household's budget constraint yields the aggregate resource constraint in Moscarini and Postel-Vinay (2019).

# 4 Empirical Strategy

In section 4.1 we discuss the calibration strategy, and in Section 4.2 the impulse responses. In Section 4.3, we show how to use the observed UE and EE flow rates to measure the stock of bad matches $l^0_{b,t}$, the on-the-job search rate, and our indicator of inter-firm wage competition $1 - \Sigma_t$, which we introduced in Section 3.6. In Section 4.4, we measure the quantitative link between our indicator of wage competition and inflation. We evaluate the model's ability to explain actual inflation in Section 4.5. In this section, we also compare how our estimate of the rate of on-the-job search based on UE and EE flow rates compares with direct survey evidence and discuss the limitations of our empirical approach.

## 4.1 Calibration

We calibrate the steady state of the model to the U.S. economy at monthly frequencies. To do so, we assume a Cobb-Douglas matching function $M_t = \phi_0 \left[ u_{0,t} + s_t (1 - u_{0,t}) \right]^{1-\psi} v_t^{\psi}$, where $\psi \in (0,1)$ is an elasticity parameter and $\phi_0 > 0$ is a scale factor.

The calibration of the steady state requires assigning values to the following 11 parameter values: $\beta$, $\phi_0$, $\delta$, $y_b$, $y_g$, $\nu$, $b$, $\xi_g$, $c$, $c^f$ and $s$. We set the discount factor $\beta$ to match an annual real interest rate of 1.5%.[15] We normalize $\theta$ to unity, which allows us to pin down the scale factor

[15] We take the average of the median projections based on the *Summary of Economic Projections* (SEP) from the FOMC meeting of May 2012—the first meeting after which the projections were released—through the meeting of December 2019.

| Calibration | | | |
|---|---|---|---|
| Parameters | Description | Value | Target/source |
| *Parameters that affect the steady state* | | | |
| $\beta$ | Discount factor | 0.9987 | Real rate 1.5%. (FOMC SEP) |
| $\phi_0$ | Scale parameter matching fn | 0.3284 | Job finding rate - Shimer (2005) |
| $\delta$ | Job separation rate | 0.0200 | Unemployment rate ($100u_{0,t}$) 5.5% |
| $y_b$ | Productivity bad matches | 1.0000 | Normalization |
| $y_g$ | Productivity good matches | 1.0800 | Faberman et al. (2019) |
| $\nu$ | Prob. of job switching if indifferent | 0.5000 | |
| $b$ | Utility of leisure | 0.8082 | Calibrated |
| $c$ | Flow cost of vacancy | 0.0124 | Calibrated |
| $c^f$ | Fixed cost of hiring | 0.4958 | Calibrated |
| $s$ | On-the-job search rate | 0.2598 | Calibrated |
| $\xi_g$ | Probability draw good match | 0.2800 | Calibrated |
| $\chi$ | Markup parameter | 1.2000 | 20% markup |
| $\varkappa$ | Scale param. Kimball aggregator | 10.0000 | Smets and Wouters (2007) |
| $\zeta$ | Calvo price parameter | 0.9250 | Quarterly probability is 80% |
| $\Pi$ | Steady-state gross inflation rate | 1.0017 | Net inflation rate of 2% p.a. |
| $\rho_r$ | Taylor rule smoothing parameter | 0.8500 | Conventional |
| $\phi_\pi$ | Taylor rule response to inflation | 1.8000 | Conventional |
| $\phi_y$ | Taylor rule response to output | 0.2500 | Conventional |
| $\psi$ | Elasticity of matching function | 0.5000 | Moscarini and Postel-Vinay (2018) |
| $\rho_\mu$ | Autocorrel. preference shock | 0.8000 | Fixed |
| $100\sigma_\mu$ | St. dev. preference shock | 0.5883 | Volatility of the unempl. rate |
| $\rho_S$ | Autocorrel. job search rate | 0.9157 | Maximum likelihood estimation |
| $100\sigma_S$ | St. dev. of job search rate shocks | 2.5510 | Maximum likelihood estimation |
| Variable | Description | Value | Target/source |
| *Steady-state calibration targets* | | | |
| $\frac{c}{\varpi}/c^f$ | Ratio of variable to fixed cost | 0.0780 | Silva and Toledo (2009) |
| $EE \equiv \frac{s\phi\left[l_b^0\left(\xi_b\nu+\xi_g\right)+l_g^0\xi_g\nu\right]}{l_b^0+l_g^0}$ | EE transition rate | 2.58 | Pre-Great Recession EE rate (%) |
| $\theta$ | Labor market tightness | 1.000 | Normalization |
| $\frac{l_g^0}{l_g^0+l_b^0}$ | Employment share in good jobs | 0.6800 | Employment share at top 10% firms |
| $\frac{\left(v_t c+c^f\phi_t[u_{0,t}+s_t(1-u_{0,t})]\right)/H}{\varphi}$ | Hiring costs over wages | 0.6000 | Hiring costs equal 2 weeks of wages |

Table 1: Calibrated values for model parameters. Notes: FOMC SEP stands for the Federal Open Market Committee's Summary of Economic Projections. EE stands for employment-to-employment.

$\phi_0$, so as to match a job finding rate of about 33 percent.[16] The job separation rate $\delta = 0.02$ is implied by the Beveridge curve, under the assumption of a steady-state rate of unemployment of 5.5%. The productivity of a bad match is normalized to one, and the productivity in a good match is set to be 8% higher. Our targeted wage differential is in line with evidence from Faberman et al. (2019) based on the *Survey of Consumer Expectations*, which shows that residualized wage gains associated with job switching are about 8%.[17] Finally, we set the probability that workers accept an equally valuable outside offer to be $\nu = 0.5$. This value is

[16]This is the average of the job finding rate computed following Shimer (2005) over a the period January 1993-December 2018.

[17]We noticed that higher values would violate the incentive compatibility constraint, which requires that the surplus of bad matches should be positive both in steady state and in all periods of the sample used to run the empirical exercise of Section 4.5.

large enough to allow the model to match the average EE flow rate in the U.S. economy. In Appendix H, we show that perturbing the value of $\nu$ does not materially affect our results.

This leaves us with five parameters to calibrate: the parameter governing the utility of leisure $b$, the probability of drawing a good match $\xi_g$, the flow cost of advertising a vacancy $c$, the fixed cost of hiring $c^f$, and the parameter governing search intensity $s$. These are calibrated in order to match the following: *(i)* A value of expected hiring costs, including both the variable and fixed cost component, equal to two weeks of wages.[18] *(ii)* A fraction of good jobs in steady state equal to 67%, which is the share of employment for the top 10% U.S. firms by employment size in the year 2000. *(iii)* A normalized value of labor market tightness equal to one. *(iv)* A ratio of total variable costs of hiring to fixed costs $\frac{c}{\varpi}/c^f$ equal to 0.078. This value is the ratio of pre-match recruiting, screening, and interviewing costs to post-match training costs in the U.S., following Silva and Toledo (2009).[19] *(v)* A monthly job-to-job transition rate of 2.5841%, which is the average EE rate (spliced using the quit rate as explained in Section 4.3) measured in the pre-Great Recession sample. We note that the value of the parameter $s$ implied by the calibration, 0.2598, is very close to the value of 0.22, which corresponds to the fraction of U.S. workers who engage in on-the-job search every month, as measured using survey data by Faberman et al. (2019).

The calibration of the probability of drawing a good match, $\xi_g$, relies on the empirical strategy in Moscarini and Postel-Vinay (2016), who exploit the correlation between firm size and productivity by assuming that employed workers climb the ladder when moving to larger firms. In Appendix H we show that our main results are not affected by reasonable variations in the probability of meeting a good match $\xi_g$.

We set the smoothing coefficient of the Taylor rule to the value of 0.85, which corresponds to a coefficient of around 0.65 in quarterly space, and the response parameters to inflation and output to the values of 1.8 and 0.25, respectively. $\chi$ is set to equal 1.2, which implies a 20% price markup (Ravn and Sterk, 2017). The steady-state gross rate of inflation is set to equal 1.0016, which implies a central bank's annual inflation target $\Pi^*$ of 2%. Finally, we set the elasticity of vacancies in the matching function $\psi$ to equal 0.5, based on estimates that account for workers searching on the job (Moscarini and Postel-Vinay, 2018).

The slope of the Phillips curve is determined by the scale parameter of the Kimball aggregator $\varkappa$ and the Calvo parameter $\zeta$, which jointly determine the degree of price stickiness. The former is set to 10 as in Smets and Wouters (2007) and the latter to 0.925, which in quarterly frequency implies a probability of not being able to reoptimize prices equal to 0.8. We set

[18] The average wage is proxied by the price of the labor service $\varphi$. If we set a target higher than two weeks of wages, the deflationary pressures predicted by the model in the past decade become slightly stronger, strengthening the ability of our model to explain the missing inflation.

[19] Silva and Toledo (2009) indicate in Table 1 (p.80), that the average pre-match recruiting cost costs is \$105.1, while the average post-match training cost amounts to \$1,359.4.

the Calvo parameter so that the implied slope of the Phillips curve allows the model to fit inflation at the beginning of the post-Great Recession recovery (2009–2011), following the same approach of Section 2. The Kimball aggregator allows us to obtain the targeted value for the slope of the Phillips curve without assuming an implausibly large degree of price stickiness.[20]

As we show in Section 4.3, we can use the observed UE and EE flow rates in combination with a subset of model equations to obtain the time series of the on-the-job search rate. This series can be retrieved from the data with no need to solve the model. To pin down this series, we just have to take a stand on a few steady-state parameters (e.g., the steady-state job finding rate $\phi$ and the separation rate $\delta$), which we calibrate using the values shown in Table 1. We use this series to estimate the persistence parameter $\rho_S$ and the standard deviation $\sigma_S$ via maximum likelihood. As for the preference shock, we set the autocorrelation parameter $\rho_\mu$ to 0.80 and then we calibrate the standard deviation $\sigma_\mu$ so that the model can match the volatility of the observed unemployment rate in the data (April 1990 through December 2018).[21]

The model is log-linearized around its steady-state equilibrium.[22] However, the zero lower bound introduces a nonlinearity that prevents us from solving the model with standard solution methods (e.g., Chris Sims' *Gensys*). To overcome this problem, we impose the zero lower bound by adding news shocks following the method described by Holden (2011), Holden and Paetz (2012), Bodenstein, Guerrieri, and Gust (2013), and Lindé et al. (2016 Section 5). This approach allows us to solve the model with standard solution techniques while allowing temporarily non-linear dynamics induced by the ZLB constraint. More details are provided in Appendix I.

## 4.2 Impulse Responses

Figure 3 shows that a fall in the rate at which workers search on the job lowers the intensity of wage competition $1-\Sigma$, where $\Sigma$ is defined in equation (16). This happens as the fraction of job seekers that are unemployed increases. Moreover, a fall in the rate of on-the-job search reduces transitions into good jobs and therefore increases the share of workers employed in bad jobs.

[20] While the slope of the Phillips curve affects the magnitude of inflation predicted by the model, it has negligible effects on the point in time when the model predicts inflation to rise above its long-run level. Therefore, the ability of our model to explain the missing inflation of the last decade is not affected by varying the slope of the Phillips curve.

[21] The value of the autocorrelation parameter is a bit lower than what is needed to fit the persistence in the U.S. civilian unemployment rate. However, a persistence higher than 0.8 would make this shock propagate as a supply shock moving the unemployment rate and inflation in the same direction. Because the shock to the on-the-job search rate propagates as a supply shock, the model would lack a demand shock to explain periods in which inflation and the unemployment rate negatively comove. Intuitively, if a negative preference shock is very persistent, the fall in vacancy creation becomes so large that it generates a sharp and prolonged contraction in the supply of the service, which in turn implies a persistent increase in its price, i.e., the real marginal cost $\varphi_t$. Moreover, the rise in current and future expected marginal costs entails a rise in the rate of inflation, together with a contraction in aggregate production.

[22] Rates and shares are linearized; all the other variables are log-linearized.

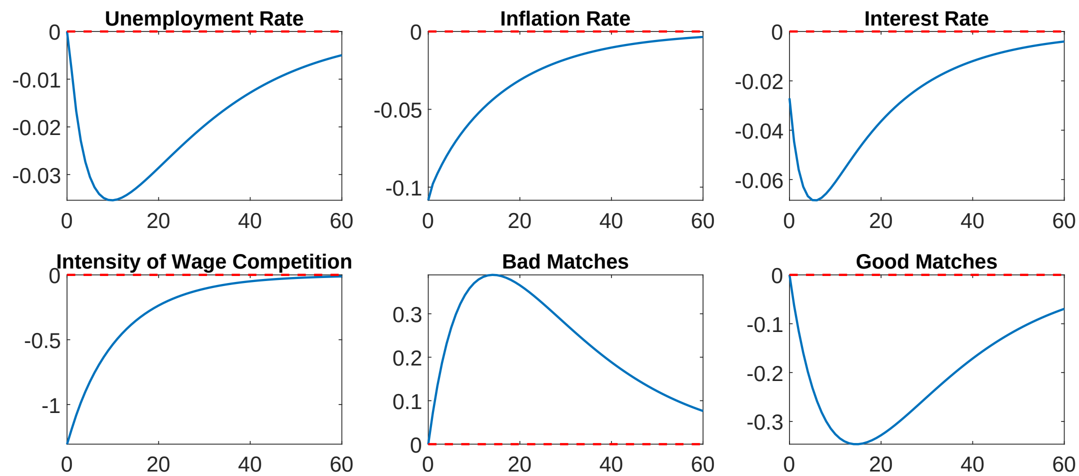


Figure 3: Impulse responses to a shock that lowers the on-the-job search rate by one standard deviation. The unemployment rate, the fraction of unemployed job seekers, and the shares of good and bad matches are measured at the beginning of the period, consistently with the definition of labor market slack $\Sigma_t$ in equation (16). Units: percentage points. Inflation and interest rate are expressed in annualized rates.

The increase in both the share of unemployed job seekers, and mismatched workers implies that in expectation, producing labor service becomes cheaper for an entrant firm. The fall in the average cost of the labor service is then passed through to lower price inflation. In response to that, the central bank cuts the interest rate, stimulating aggregate demand and reducing unemployment. Moreover, attracted by the expectation of cheaper labor, more vacancies are created, which also contributes to lowering the unemployment rate.[23] Interestingly, a negative shock to the rate of on-the-job search can generate simultaneously a persistent rise in output, together with a fall in unemployment, inflation, and productivity. Incidentally, these patterns seem to accord well with the dynamics that have characterized the U.S. economy towards the end of the second decade.

A negative preference shock raises the unemployment rate and cyclical misallocation while it lowers inflation. Because the propagation of this shock is rather standard, we show the impulse responses in Appendix J.

## 4.3 The On-the-Job Search Rate in the Macro Data

We show that for a given value of bad and good matches at the beginning of our sample period (i.e., in April 1990), observing UE and EE rates pins down the entire time series of the on-the-job search rate $s_t$, as well as the time series of bad matches $l^0_{b,t+1}$ and good matches $l^0_{g,t+1}$. The exact identification of these variables comes from a set of the model's equations and does not

[23] Note that the fall in the unemployment rate, in isolation, contributes to lowering the probability for an entrant firm to meet an unemployed worker and hence causes wage competition to become more intense. In equilibrium though, this effect is dominated by the fall in the rate of on-the-job search, which operates in the opposite direction, raising the fraction of unemployed job seekers.

require solving the model.[24]

Consider the measurement equation for the EE flow rate ($EE_t$):

$$EE_t \equiv \frac{s_t \phi\left(\theta_t\right)\left[l^0_{b,t}\left(\xi_b \nu + \xi_g\right) + l^0_{g,t}\xi_g \nu\right]}{l^0_{b,t} + l^0_{g,t}}. \tag{20}$$

This rate is the ratio of how many workers employed at the beginning of the period switch jobs (the EE flows) to the total numbers of workers employed at the beginning of the period. Consistently with our model, the EE flows are given by the sum of all the workers who find a better match and the fraction $\nu$ of those workers who move into an equally valuable match.

Linearizing equation (20), and using the symbol $\tilde{}$ to denote linearized variables yields the following equation:

$$\begin{aligned} \tilde{s}_t &= \frac{s}{EE}\widetilde{EE}_t - \frac{s}{\phi}\tilde{\phi}_t - \frac{s}{\nu\left(l^0_b + l^0_g\right)}\left[\frac{s\phi\left[\left(\xi_b + \nu\xi_g\right)\right]}{EE} - 1\right]\tilde{l}^0_{b,t} \\ & -\frac{s}{\nu\left(l^0_b + l^0_g\right)}\left[\frac{s\phi\xi_g}{EE} - 1\right]\tilde{l}^0_{g,t}, \end{aligned} \tag{21}$$

which expresses the on-the-job search rate $\tilde{s}_t$ as a function of the observed EE flow rate $\widetilde{EE}_t$, the UE rate $\tilde{\phi}_t$, and the measure of good and bad matches at the beginning of period $t$ ($\tilde{l}^0_{b,t}$ and $\tilde{l}^0_{g,t}$). Because $\tilde{l}^0_{b,t}$ and $\tilde{l}^0_{g,t}$ are predetermined at time $t$, this equation allows us to exactly measure $\tilde{s}_t$ consistently with the series for $\tilde{\phi}_t$ and $\widetilde{EE}_t$.

With the rates $\tilde{\phi}_t$ and $\tilde{s}_t$ at hand, we can pin down the fraction of bad and good matches in the next period $t+1$, using the linearized laws of motion for low- and high-quality matches in (17) and (18):

$$\tilde{l}^0_{b,t+1} = -\left(1-\delta\right)\left\{\phi\xi_g l^0_b \tilde{s}_t + \left[s\xi_g l^0_b - \xi_b u_0\right]\tilde{\phi}_t\right\} + \left(1-\delta\right)\left\{\left[1 - s\phi\xi_g\right]\tilde{l}^0_{b,t} + \phi\xi_b\tilde{u}_{0,t}\right\}, \tag{22}$$

$$\tilde{l}^0_{g,t+1} = \left(1-\delta\right)\left[\tilde{l}^0_{g,t} + \phi\xi_g l^0_b\tilde{s}_t + s\phi\xi_g\tilde{l}^0_{b,t} + \phi\xi_g\tilde{u}_{0,t} + \left[s\xi_g l^0_b + \xi_g u_0\right]\tilde{\phi}_t\right]. \tag{23}$$

This requires using equations (3) to (5) to solve for the linearized unemployment rate $\tilde{u}_{0,t} = \left(1-\phi_{t-1}\right)\left(1-\delta\right)\tilde{u}_{0,t-1} - u_0\left(1-\delta\right)\tilde{\phi}_{t-1}$. The value of $\tilde{u}_{0,t}$ at the beginning of the sample is initialized at its steady state value. With the knowledge of the distribution of match quality at time $t+1$, we can go back to equation (21) and obtain the on-the-job search rate in period $t+1$ ($\tilde{s}_{t+1}$). Iterating over these steps yields the whole time series for $\tilde{s}_t$, $l^0_{g,t}$ and $l^0_{b,t}$. In turn, this allows us to compute the indicator of wage competition using equation (16). It is important to notice that solving the model is not needed to pin down exactly these time series. Similarly, we

[24] We assume that the distribution of match quality is at steady state in April 1990. The results would not change if we introduced a Gaussian prior reflecting uncertainty about the initial conditions and then used the Kalman filter to optimally estimate these initial conditions.

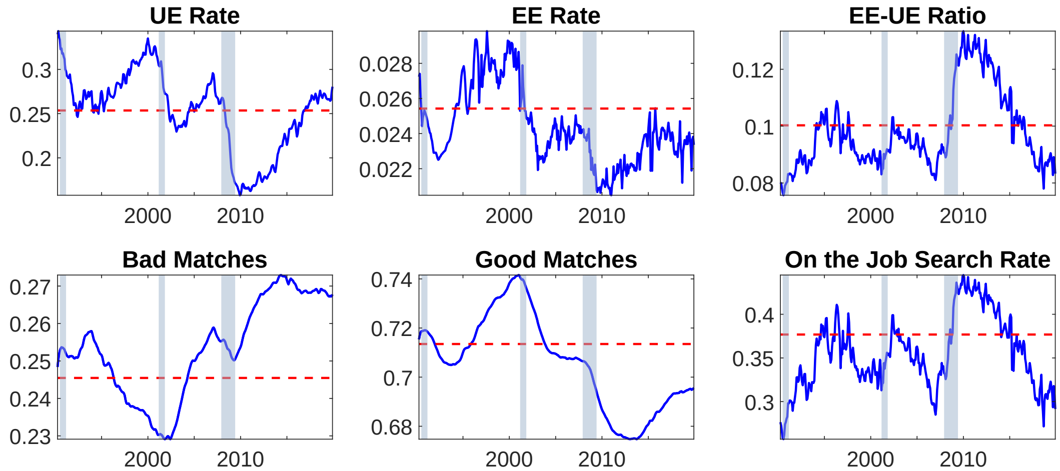


Figure 4: Labor market variables simulated from the calibrated model using the shocks that allow the model to explain exactly the observed unemployment-to-employment (UE) flow rate and the employment-to-employment (EE) flow rate. The red dashed lines denote the model's steady-state value of each simulated variable.

can estimate the persistence and standard deviation of the shock to search intensity, i.e. the parameters $\rho_s$ and $\sigma_s$, before solving the model, based on the implied series for $s_t$.

We measure UE and EE flow rates from the *Current Population Survey* (CPS). We correct the EE series as suggested by Fujita, Moscarini, and Postel-Vinay (2019) and extend it back to April 1990 by splicing this series with the quit rate measured by Davis, Faberman, and Haltiwanger (2012). While the main focus of the paper is on the period that follows the Great Recession, which is when the standard theories of inflation most significantly fail, we show the behavior of the rate of on-the-job search and our measure of bad jobs over this longer period of time.

Figure 4 shows the dynamics of the UE and EE rates observed over the period April 1990 through December 2018, together with the implied paths for the rate of on-the-job search and the stock of bad and good matches.[25] The UE rate suggests that the U.S. labor market became quite tight towards the end of the second decade; however, the dynamics of two key drivers of the indicator of wage competition in equation (16), i.e., the on-the job-search rate and the stock of bad matches, paint a different picture. After the Great Recession, the rate of on-the-job search fell to a historically low level. As a result, the stock of bad matches increased, remaining persistently elevated throughout the recovery and contributing to reduce inflationary pressures and labor productivity.[26] The prediction that bad jobs were still heightened late in the sample

[25] It is important to keep in mind that given the stylized nature of the model the uncertainty about these estimates is quite large.

[26] Haltiwanger, Hyatt, McEntarfer, and Staiger (2021) show that the net poaching rate has completely recovered by 2015. This is also the case in our model, suggesting that the finding that the net poaching rate has completely recovered by 2015 is not inconsistent with the persistent rise in the stock of bad jobs in the economy shown in Figure 4. This is because in our calibrated model the flows affect the behavior of stocks slowly over time.

period is consistent with the *Survey of Consumer Expectations*, which shows that about 30% of the workers employed in 2017—after eight years of recovery—were not fully satisfied with how their current jobs fit their experience and skills.[27]

The ratio of EE to UE flows, in the top right corner of Figure 4, falls below its steady state value towards the end of the second decade. This is a direct consequence of the fall in the rate of on-the-job search. In principle, the EE to UE rate could have been low if, absent time-variation in $s_t$, the employed workers were efficiently allocated at the top of the ladder after a decade of expansion. Distinguishing between the two hypothesis is important because of their opposite implications for inflation dynamics. Inflation rises when the employed workers are efficiently allocated (see Moscarini and Postel-Vinay, 2019), whereas it falls if the employed search less.

Eeckhout and Lindenlaub (2019) estimate the intensity of on-the-job search over the business cycle, exploiting both EE and UE flows. They assume that only workers in bad matches actively search on the job. We relax this assumption because in our model the bargaining protocol follows Postel-Vinay and Robin (2002). In this setup, workers employed in good matches too have an incentive to search, as attracting outside offers may lead to a larger surplus extraction. Under the assumption that workers in both good and bad matches search on the job, the estimated on-the-job search rate appears to lead to a more pronounced fall in expansions. As we will show at the end of Section 4.3, the behavior of the estimated rate of on-the-job search closely mimics the value computed using micro survey data in the available years.[28]

It is important to clarify that when the model's variable $s_t$ is measured from equation (21), it effectively picks up a wedge between EE and UE rates, which may as well confound other effects. For instance, while the model abstracts from the intensive margin of on-the-job search, the fall in $s_t$ measured from the macro data could potentially reflect a decline in the average number of hours spent searching. Alternatively, while the model assumes that conditional on searching, both unemployed and employed workers find jobs at the same rate $\phi(\theta_t)$, it may well be that in the data the arrival rate of job offers, conditional on searching, has diverged for these two types of job seekers, with offers becoming less frequent for the employed workers relative to the unemployed. This could be the case, for instance, if over time the employed workers had experienced a decline in the availability of suitable jobs relative to the unemployed or just faced more stringent hiring practices. A formal analysis of these wedges is provided in Appendix Q.

As we shall show at the end of Section 4.5, the estimated fall in the model's on-the-job search

[27] One of SCE questions reads as follows: "On a scale from 1 to 7, how well do you think this job fits your experience and skills?" About 30% of the respondents reported a satisfaction of 5 or less.

[28] One can be concerned that we do not allow the probability of drawing a good match $\xi_g$ to vary over time. The empirical evidence reviewed by Barlevy (2002) suggests that the quality of new matches declines in recessions and increases in expansions (see also Haltiwanger, Hyatt, McEntarfer, and Staiger (2021) for more recent evidence). Making $\xi_g$ procyclical would imply an even more countercyclical estimate of $s_t$, which in turn would imply an even bigger fall in the rate of on-the-job search towards the end of the second decade. To see why, notice that the right-hand side of equation (20) is increasing in $\xi_g$. For given EE and UE rates and a beginning-of-period distribution of match quality, an increase in $\xi_g$ would imply a lower value for $s_t$.

rate, $s_t$, is supported by survey evidence for the period of the missing inflation 2014-2019. While this constitutes an important external validation for our explanation of the missing inflation, the short sample period of the available survey leaves the question of how the on-the-job search rate behaves over the business cycles unanswered. Thus, the countercyclical dynamics that our model imputes to the on-the-job search rate in Figure 4 may well, in fact, be driven by the other wedges described in Appendix Q.

The model that we have considered so far assumes that on-the job search is exogenous. One advantage of this approach, is that it allows us to derive a closed-form indicator of wage competition and inflationary pressures, which can be measured from the UE and the EE transition rates, without solving the model. A disadvantage, is that it does not capture the incentives that workers may have to search more actively when employed in a bad match or during expansions, when getting job offers is easier. These effects may reduce the fall of the on-the-job search rate in recent years, which is key to explaining the missing inflation. In Appendix N we therefore endogenize the rate of on-the-job search, so as to properly model how the propensity of workers to search on the job differs conditional on their job quality. We show that doing so does not materially affect our conclusions on the dynamics of wage competition and on the missing inflation in the previous decade. This is because the rate of on-the-job search is effectively implied by the joint behavior of UE and EE rates, even in this richer version of the model.

### 4.4 A Useful Decomposition of Inflation

As mentioned in Section 3.6, and shown in Appendix G, under some parameter restrictions, inflation in the model can be decomposed into three labor market variables: $u_0$, $s_t$, $l^0_{b,t}$. In this section, we show that in our calibrated model, inflation can indeed be accounted for by the behavior of the three labor market variables above with a good degree of approximation. We do so by regressing the simulated series for inflation against the simulated series of the three explanatory variables and obtain:

$$\hat{\pi}_t = \underset{[-0.0181,-0.0180]}{-0.0181} \; \tilde{u}_{0,t} - \underset{[-0.073, \quad -0.0073]}{0.0073} \; \tilde{l}^0_{b,t} + \underset{[0.0029, \quad 0.0029]}{0.0029} \; \tilde{s}_t, \tag{24}$$

where the numbers in square brackets under the estimated coefficients denote 95% confidence intervals. The R-squared of the ordinary least squares (OLS) regression is 0.96, which signifies a remarkably good ability of the three labor market variables to explain inflation in the model.

This very good fit is not surprising in light of the proof—provided in Appendix G—that inflation is driven by the unemployment rate, the degree of labor misallocation, and the on-the-job search rate under some parametric restrictions. Since the calibrated model does not satisfy these parametric restrictions perfectly, the fit measured by the R-squared cannot be perfect. Nevertheless, this statistic is sufficiently close to one to believe that the three labor market

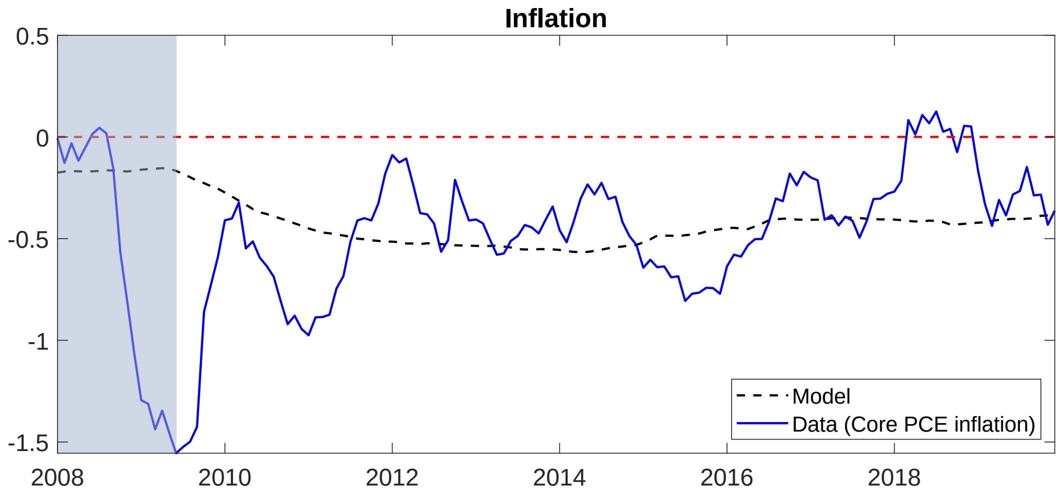


Figure 5: Inflation dynamics. The blue line denotes the U.S. annualized core PCE inflation rate in percentage deviation from the *Survey of Professional Forecasters* 10-year average expectations about the PCE inflation rate. The black dashed line denotes the model's predicted annualized inflation rate in percentage deviation from its long-run steady-state value.

variables jointly capture the bulk of the variations of inflation in the calibrated model. This result is very useful because it will allow us to show how these three labor market variables contributed to moving inflation in the past decade and which one played the most important role in explaining the missing inflation. This is the topic of the next section.

In addition, a model-consistent measure of inflationary pressures can be estimated without the need to solve the model. This can be done in two steps. First, data on UE and EE flow rates can be used to estimate the three labor market variable on the right-hand side of equation (24) as shown in Section 4.3. Second, endowed with the time series of unemployment, the share of low-productivity jobs, and the on-the-job search rate, we can use equation (24) to retrieve our model's prediction of inflation.

## 4.5 The Missing Inflation Puzzle Explained

We want to evaluate the ability of the model to explain the missing inflation during the recovery that followed the Great Recession. We rely on the same data set used to measure the on-the-job search rate in Section 4.3, which comprises only the UE and EE flow rates. We then use our linearized model to retrieve the series of two shocks that make the model explain exactly these data. We then feed the model with these shocks to simulate the inflation rate predicted by the model in the last decade.[29]

Figure 5 compares the year-over-year core PCE inflation rate in the data (the blue line) with the rate of inflation simulated from the model (the black dashed line). The key takeaway

[29] As before, we assume that the economy is in steady state at the beginning of the sample period. Different assumptions on the initial conditions would not affect our results because the sample period begins in April 1990 and the analysis focuses on a briefer period that starts several years later (specifically, in January 2011).

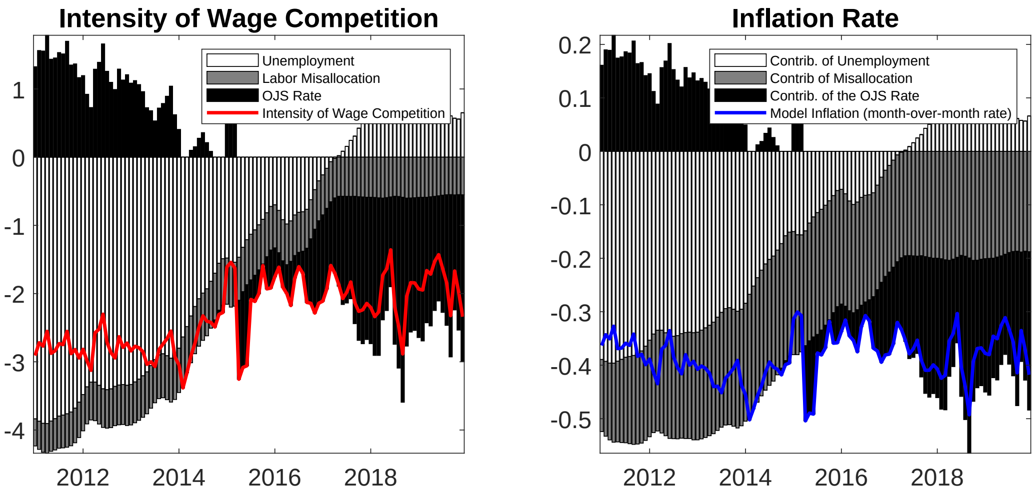


Figure 6: The intensity of interfirm wage competition and inflation and their decomposition. Left panel: The red solid line is the estimated intensity of interfirm wage competition in percentage deviation from its steady-state value. The bars denote the contribution of the unemployment rate (white bars), labor misallocation (gray bars) and the on-the-job (OJS) search rate (black bars) based on the linearization of equation (16). Right panel: The blue solid line is the model's predicted inflation rate in percentage deviation from its steady-state value. The bars denote the contribution of the unemployment rate (white bars), labor misallocation (gray bars) and the on-the-job (OJS) search rate (black bars) based on the estimated equation (24).

is that the model can account for the rate of inflation remaining persistently below its long run average throughout the second decade. The fit of the high-frequency behavior of inflation cannot of course be perfect, given that the empirical exercise only relies on two labor market variables as observables and two shocks.

In order to shed light on the labor-market forces driving inflation dynamics, the left panel of Figure 6 plots the linearized indicator of inter-firm wage competition, based on equation (16), and decomposes its fluctuations into the contributions of its three components. Similarly, the right panel of the same figure decomposes the rate of inflation on the same three labor market variables using the decomposition in equation (24).[30] A first result to take away from the figure, comparing the two panels, is that the indicator of wage competition correlates almost perfectly with inflation in the model, and that the labor market forces that drive the two are also very similar. A second result is that the indicator of wage competition can successfully explain the low inflation observed during the previous decade, unlike the traditional measures of slack analyzed in Figure 2.

The decompositions of both the indicator of wage competition and inflation indicate that at the beginning of the recovery, inflation was low primarily because of the persistent surge in the unemployment rate caused by the Great Recession, as illustrated by the white bars. Towards the latest part of the decade, further improvements in aggregate labor market conditions quickly lowered the share of unemployed job seekers, causing the unemployment rate to reverse the sign of its contribution to inflation. However, from about 2015 and onwards, the on-the-job search

[30] We linearize the indicator of wage competition to allow both for a comparison with the behavior of model inflation, which is also linearized, and for a decomposition of the driving forces.

rate declined rapidly, putting downward pressures on inflation (the black bars) and dominating the effects of the unemployment rate (the white bars) towards the latest part of the sample.

The role played by the incidence of bad matches is also very interesting (the gray bars in the left panel of Figure 6). Bad matches have always contributed to keeping inflation below its long-run level. According to the model's results, in the earlier part of the period of interest, after the unusually severe recession, a large fraction of unemployed workers took a first step onto the ladder, raising the stock of bad jobs. Later in the recovery, as the on-the-job search rate declined sharply, the speed at which workers moved to better jobs fell, exacerbating labor misallocation and keeping the intensity of wage competition low. The gray bars precisely highlight the role played by the cyclical match composition of the employment pool in explaining the missing inflation.

The focus of this section has been on the ability of the model to explain the behavior of inflation in the period that followed the Great Recession. In Appendix K we compare how the indicator of interfirm wage competition $(1 - \Sigma_t)$ fares, relative to traditional measures of slack, in explaining the behavior of inflation over a longer sample, which is extended to include the 1990s and the 2000s. We show that with respect to the traditional measures of slack considered in Section 2, the indicator of wage competition achieves the lowest Root Mean Squared Error on the longer sample. We also carry out a number of other robustness exercises. Namely, we check that our results are robust to including more labor market variables in estimation, such as the vacancy filling rate, the unemployment rates, an extended series of unemployment rate that accounts for workers marginally attached to the labor force (U5), and a UE flow rate that accounts for transitions from inactivity, replicating and expanding the series introduced by Hall and Schulhofer-Wohl (2018). The results are reported in Appendix L.

**The On-the-Job Search Rate in the *Survey of Consumer Expectations*** We now look into the micro data to validate our macro-based measurement of the on-the-job search rate. To this end, we explore a new survey that is informative about the search behavior of the employed workers; it has been administered by the Federal Reserve Bank of New York as a supplement to the *Survey of Consumer Expectations* (SCE). The SCE is a monthly and nationally representative survey of about 1,300 individuals. This survey is very useful for our purpose because it directly asks employed workers whether they have been actively searching for work in the previous seven days.[31] Furthermore, this data set overcomes a typical problem characterizing other data sets (e.g., *the American Time Use Survey*) that report very small times spent in searching for a new job. In this paper, we use SCE data available from 2014 trough 2017. Even if this is admittedly a short period of time, it still covers the four years in

[31] Question JS9 of the survey asks the following: "And within the LAST 7 DAYS, about how many TOTAL hours did you spend on job search activities? Please round up to the nearest total number of hours." We drop self-employed workers when computing the on-the-job search rate from the SCE.

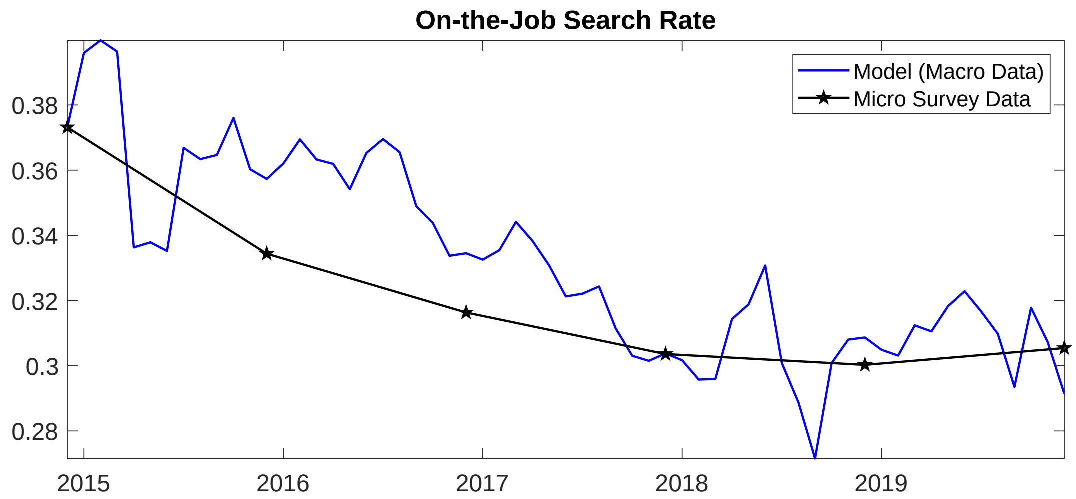


Figure 7: The estimated on-the-job search rate in the model (blue solid line) and in the *Survey of Consumer Expectations* (the red line with the star marker). In the *Survey*, the rate is computed by dividing the workers who have searched at least one hour within the last seven days by the total number of workers surveyed. The rate is conditional on those surveyees who are working for someone else. The first observation of the Survey data is rescaled so as to be the same as the corresponding on-the-job search rate estimated with the macro data

which the fall in the rate of on-the-job search predicted by our model is critical to account for the missing inflation.

Figure 7 plots the on-the-job search rate implied by the model, $s_t$, and the corresponding measure in the micro data (blue solid line and the black dashed-dotted line, respectively). The figure shows that the fall in the on-the-job search rate predicted by our model using aggregate labor market flows is strikingly close to the one measured in the micro data.

Using information on the hours of search for the employed workers in the SCE, we find that the fall in the aggregate amount of time spent searching in our model in the 2014-2019 period is entirely explained by a fall in the extensive margin in the Survey data. The effect is due to a fall in the incidence of job search among the employed—and not to a decrease in the average number of hours dedicated to search. We also looked at how the arrival rate of job offers for the employed workers varied over our sample period, relative to the arrival rate of offers for the unemployed. The evidence does not indicate a divergence in the arrival rate of offers for the employed and the unemployed.[32] Therefore, the SCE validates the decline of the on-the-job search rate predicted by our macro model over the years that are crucial to account for the missing inflation over the post-Great Recession recovery.

[32] We computed, both for the employed and the unemployed, the ratio between the total number of offers received—and not necessarily accepted—and the aggregate total number of hours spent searching.

# 5 The Role of Labor Costs

Weak interfirm wage competition lies at the hearth of our explanation for the missing inflation puzzle. Figure 6 suggests that the reason why price inflation has become increasingly disconnected from unemployment is because wage inflation has also become increasingly disconnected. This is due both to the direct deflationary effect that the fall in on-the-job search produces on wage inflation, and its indirect effect that operates by prolonging the increase in cyclical misallocation. Previous work by Coibion and Gorodnichenko (2015) has documented that up until the end of 2013 there was evidence of a disconnection of inflation from the unemployment rate but no evidence of a disconnection of wage inflation from the unemployment rate. If this result were confirmed for the later years, it would clearly undermine our explanation of the missing inflation, which rests on a decline of interfirm wage competition and hence shallow wage dynamics in the second half of the past decade.

To investigate this issue, we extend the analysis of Coibion and Gorodnichenko (2015) through the end of 2019. Specifically, we regress quarterly wage inflation ($\pi_t^w$) net of expectations of wage growth on unemployment rates:

$$\pi_t^w - E\pi_{t+1}^w = c + \beta u_t + error_t. \tag{25}$$

The coefficient $\beta$ can be interpreted as the slope of the wage Phillips curve. A flat wage Phillips curve is an indication of disconnection between wage inflation and unemployment. According to our story, the missing inflation is explained by a decline in the on-the-job search rate that lowered wage pressures notwithstanding the remarkable decline of the unemployment rate during the post-Great Recession recovery. Therefore, our explanation calls for a flattening of the estimated wage Phillips curve—that is, the estimated $\beta$ in the past decade should fall relative to its estimated value in the earlier period.

First, we assume backward looking expectations, and compute expected inflation as an average of wage inflation in the previous four quarters. Panels A and B of Figure 8 plot the results obtained measuring wages as average hourly earnings of manufacturing workers, and nonfarm business sector compensation per hour, respectively. The results indicate that while the slope of the Phillips curve has remained rather stable until the Great Recession, there has been a noticeable flattening of the curve in the following period, which includes data all the way up to 2019Q4. To assess whether this result is robust to assumptions about expectations of wage inflation, we consider weekly manufacturing earnings as a third measure of compensation. This measure is useful because the Livingstone Survey of Forecasters provides real-time forecasts of wage inflation six and twelve months ahead. Panel C of Figure 8 plots semiannual deviations of wage inflation from expected wage inflation against semiannual unemployment from 1960S1 to 2019S2. Once again, the results indicate that the wage Phillips curve has been stable until

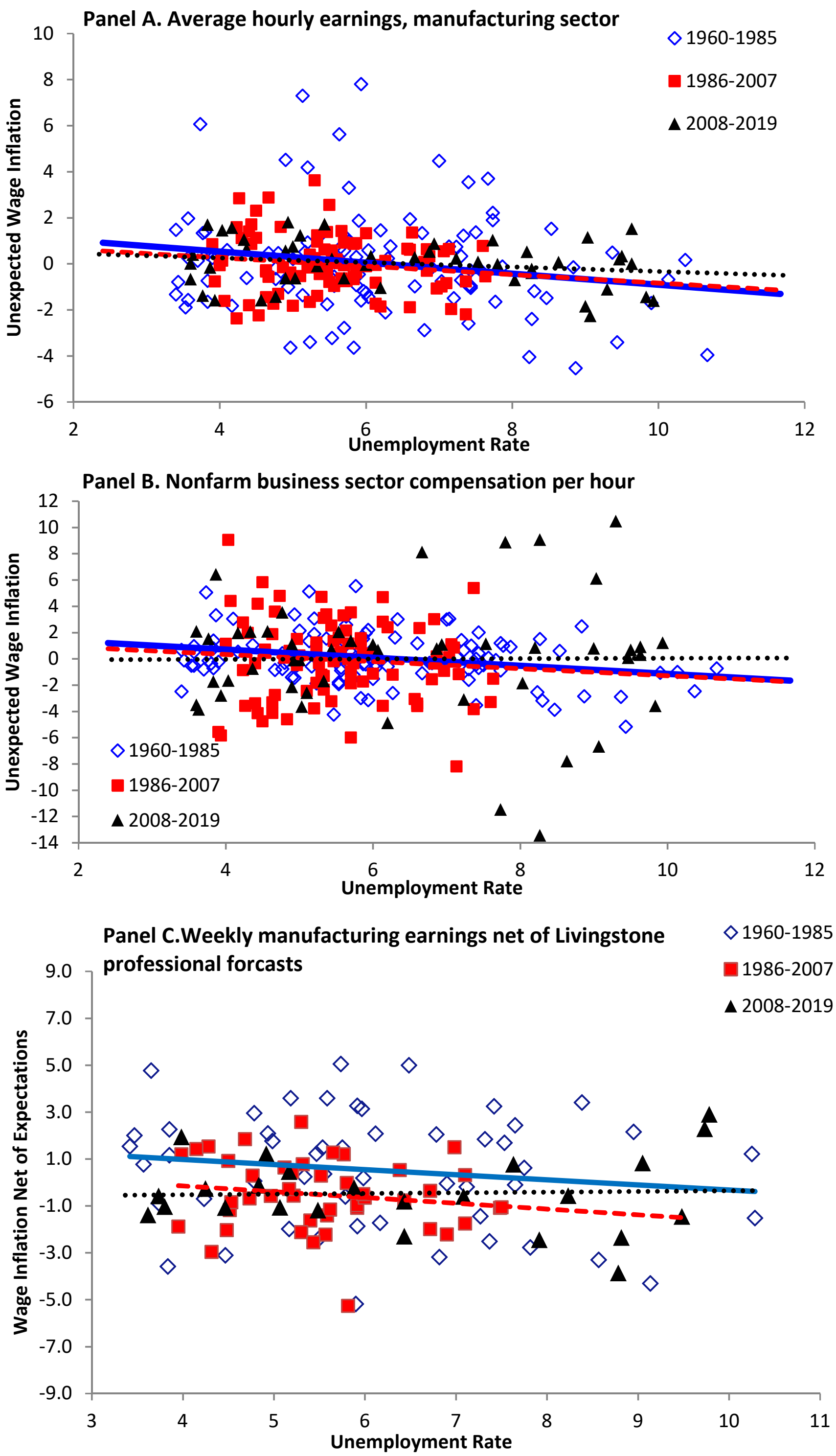


Figure 8: Each panel plots a measure of wage growth net of wage growth expectations against the unemployment rate, using different measures of wages in each panel. Panels A and B use backward-looking measures of wage expectations. Panel C uses forecasts from the Livingston Survey of Professional Forecasters. The solid blue lines are average slopes for 1960-1985, the dashed red lines are average slopes for 1986-2007 and the dotted black lines are average slopes for 2008-2019. Panel C uses semiannual data.

the pre-Great Recession sample, and then flattened in the subsequent decade. Comparing our results with those in Coibion and Gorodnichenko (2015) suggests that the stability of the wage Phillips curve must have broken around the second half of the previous decade, in line with the story that emerges from Figure 6.

We note two key takeaways from the analysis shown in this section. First, up until 2013 our model would not have predicted a disconnection between wage dynamics and unemployment in line with the finding in Coibion and Gorodnichenko (2015). To see this, note the large and dominant negative contribution of the unemployment rate on interfirm wage competition over the early years of the post-Great Recession recovery in the left panel of Figure 6. Second, when we extend the analysis of Coibion and Gorodnichenko (2015) to this later period, we find a clear disconnection between wage growth and the unemployment rate, consistently with the prediction of our model. Our model explains this disconnection in the second half of the past decade with the declining on-the-job search rate, which dominates the effects of the falling unemployment rate. See the left panel of Figure 6.

# 6 Post-Pandemic Job-to-Job Transitions and Inflation

In this section, we turn to the question as to whether our model can explain some of the considerable rise in inflation observed during the pandemic period. To address this question, we extend the model of Section 3 to allow for the participation rate to change over time. This is a sensible modification, given the steep fall in the labor force observed around the beginning of the pandemic. We take the model to the data using a procedure that is based on the one used in the pre-pandemic analysis and described in Section 4.3. Unlike in the pre-pandemic period, we now also make use of data on the unemployment rate and labor force participation rates to retrieve both the rate of on-the-job search and the evolution of match quality. To do so, we introduce shocks to the separation rate and to the participation rate. Furthermore, given the massive shock that hit the labor market in April 2020, we do not linearize the model equations as we did in Section 4.3. Appendix M provide more details about how the model is taken to the data in the pandemic period.

The right panel of Figure 9 shows the rate of on-the-job search estimated by the model during the pandemic period. The results indicate that the rate of on-the-job search has initially fallen at the very beginning of the pandemic, subsequently increased steadily since July 2020, peaking in June of the next year. The rate is estimated to fall below its pre-pandemic level (the horizontal dashed red line) toward the end of 2021. The pattern of the on-the-job search rate broadly mimics that of the EE to UE ratio, depicted in the left panel of Figure 9. So according to our estimate, this ratio did not increase in 2020-2021 because labor misallocation increased —which would be deflationary— but because of an increase in the rate of on-the-job search,

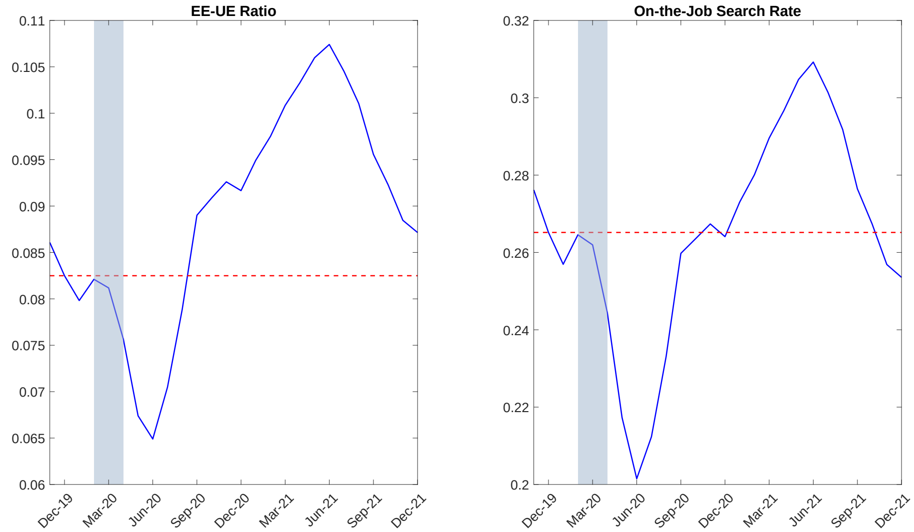


Figure 9: Dynamics of key labor market during the pandemic. Left panel: the EE-UE ratio observed in the CPS data. The EE flow rate is corrected as in Fujita, Moscarini, and Postel-Vinay (2019). Right panel: the estimated rate of on-the-job search. The horizontal red dashed lines denote the value of the two variables in December 2019.

which is inflationary.

Figure 10 plots the intensity of wage competition, $1 - \Sigma_t$, exploiting the non-linear equation (16), and indexing the indicator to 100 in June 2020 (the red line). Wage competition is relatively low at the beginning of the pandemic, reflecting the high rate of unemployment, and then rises as unemployment falls. Throughout the years 2020 and 2021, the increase in the rate of on-the-job search contributes positively to increase wage competition. This is shown by the red line with dots in the left panel of Figure 10, which plots the evolution of our indicator of wage competition, under the assumption that the rate of on-the-job search is fixed at its value in June 2020. The red line lies below the blue one, indicating that inflationary pressures would have been lower if the rate of on-the-job search had not increased over the period.

In the right panel of the same Figure, we show core PCE inflation in the data (the black solid line) and the inflation due to the shock to the on-the-job search rate estimated from June 2020 through December 2021 (the red bars). The red bars show that the increase in the rate of on-the-job search has contributed to about a 1 percentage point rise in price inflation for most of 2021. Interestingly, even though the estimated on-the-job search rate—shown in the right panel of Figure 9—declined substantially in the summer and fall of 2021, the model still predicts sizably positive inflationary effects through the end of the sample. These highly persistent effects on inflation are due to the job ladder mechanism: the robust increase in the on-the-job search rate from July 2020 through June 2021 has improved the allocation of labor, causing a persistent increase in the intensity of interfirm wage competition.

The results shown in this section come with some important caveats. First, it is well-known that the pandemic had large and persistent effects on the participation of the U.S.

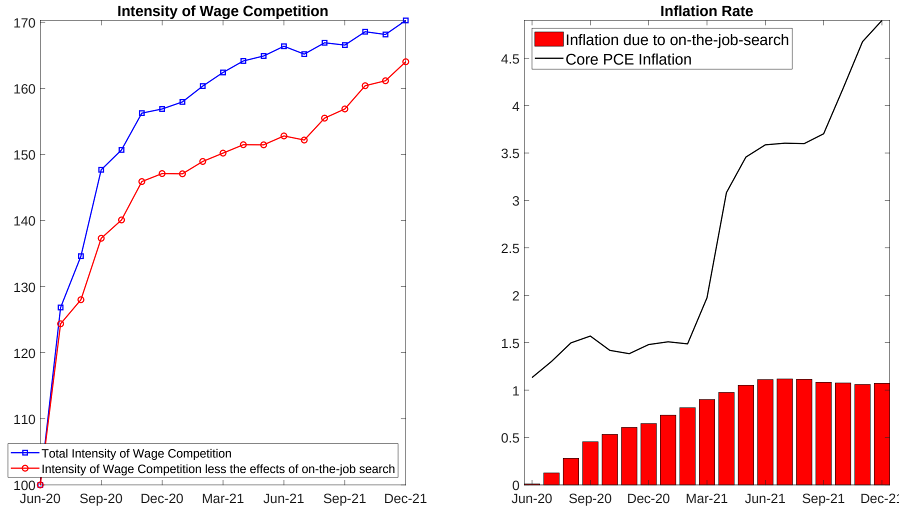


Figure 10: Wage competition among firms and inflationary pressures. Left panel: the blue line with square markers denotes the estimated intensity of interfirm wage competition. The red line with circle markers denotes the counterfactual intensity of interfirm wage competition if none of the estimated shocks to the rate of on-the-job search had occurred from June 2020 through December 2021. Right panel: The black solid line denotes the U.S. core PCE inflation (percent change from year ago). The red bars denote the contribution of the shocks to the rate of on-the-job search to model's inflation estimated rom June 2020 through December 2021.

labor force. However, we do not have epidemiologic dynamics in our model nor endogenous participation. Second, in response to the pandemic the U.S. government enacted legislation aimed at increasing the duration and the generosity of unemployment benefits. This legislation may have discouraged unemployed workers to accept job offer, artificially lowering the UE flow rate. Since our model does not feature unemployment benefits, the lower UE rate is interpreted by our model as evidence that the employed workers were relatively more successfully at finding jobs during the pandemic, which is in turn explained by an increase in the estimated on-the-job search rate. This shortcoming may have led us to overestimate somewhat the increase in the on-the-job search rate during the pandemic. Consequently, the estimated inflationary effects of "the great resignation" should be regarded as an upper bound.

# 7 Concluding Remarks

We showed that traditional measures of labor market slack fail to explain why inflation has remained subdued throughout the post-Great Recession recovery. We introduced a model with the job ladder in which the fraction of workers searching on the job influences labor market slack by affecting the degree of interfirm wage competition to hire employed workers. We found that the model explains the missing inflation of the past decade with the fall in the rate of on-the-job search and the associated weakening of wage competition among firms. Finally, we verified that when the on-the-job search rate is identified at micro levels using survey data, a similar behavior in this rate is detected from 2014 through 2019. During the period 2020-2021, dubbed the "great resignation", the indicator of interfirm wage competition rose considerably, contributing to raising inflation by 1 percentage point for most of 2021.

Yet, it is important to keep in mind that our model is stylized and the point estimate of any model's variables is subject to a great deal of uncertainty. Only further research on job-ladder models and on the drivers of workers' decisions to search on the job can tell us more about the potency of the mechanism studied in this paper.

Our study opens venues for future research on the drivers of workers' propensity to search on the job and on the appropriate stabilization policies in the presence of interfirm competition for the employed. For instance, an important question to explore is whether monetary policy, whose primary goal is to stabilize inflation, has any significant effect on the search behavior of the employed. While the empirical literature has made important progress in understanding how monetary impulses affect labor supply mobility, very little is known about the effectiveness of monetary stimuli in incentivizing workers to search on the job.

# References


Abraham, G. K., C. J. Haltiwanger, and L. Rendell (2020): "How Tight is the Labor Market?," *Brookings Papers on Economic Activity*, Spring, 97–165.

Atkeson, A., and L. E. Ohanian (2001): "Are Phillips curves useful for forecasting inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), 2–11.

Ball, L., and S. Mazumder (2019): "A Phillips Curve with Anchored Expectations and Short-Term Unemployment," *Journal of Money, Credit and Banking*, 51(1), 111–137.

Bodenstein, M., L. Guerrieri, and C. J. Gust (2013): "Oil shocks and the zero bound on nominal interest rates," *Journal of International Money and Finance*, 32, 941–967.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

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

Coibion, O., and Y. Gorodnichenko (2015): "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, 7(1), 197–232.

Coibion, O., Y. Gorodnichenko, and M. Ulate (2019): "Is Inflation Just Around the Corner? The Phillips Curve and Global Inflationary Pressures," *AEA Papers and Proceedings*, 109, 465–469.

Davis, S. J., R. J. Faberman, and J. Haltiwanger (2012): "Labor market flows in the cross section and over time," *Journal of Monetary Economics*, 59(1), 1 – 18.

Davis, S. J., R. J. Faberman, and J. C. Haltiwanger (2013): "The Establishment-Level Behavior of Vacancies and Hiring," *The Quarterly Journal of Economics*, 128(2), 581–622.

Del Negro, M., M. Lenza, G. Primiceri, and A. Tambalotti (2020): "What's up with the Phillips Curve," *Brookings Papers on Economic Activity*, Spring, 301–357.

Dotsey, M., and R. G. King (2005): "Implications of state-dependent pricing for dynamic macroeconomic models," *Journal of Monetary Economics*, 52(1), 213–242.

Eeckhout, J., and I. Lindenlaub (2019): "Unemployment Cycles," *American Economic Journal: Macroeconomics*, 11(4), 175–234.

Faberman, R. J., A. I. Mueller, A. Sahin, and G. Topa (2017): "Job Search Behavior among the Employed and Non-Employed," IZA Discussion Papers 10960, Institute of Labor Economics (IZA).

Fernald, J. G. (2016): "Reassessing Longer-Run U.S. Growth: How Low?," Working Paper 2016-18, Federal Reserve Bank of San Francisco.

Fujita, S., G. Moscarini, and F. Postel-Vinay (2019): "Measuring Employer-to-Employer Reallocation," Working paper.

Galí, J., and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, pp. 195–222.

Galí, J., F. Smets, and R. Wouters (2011): "Unemployment in an Estimated New Keynesian Model," in *NBER Macroeconomics Annual 2011, Volume 26*, NBER Chapters, pp. 329–360. National Bureau of Economic Research, Inc.

Gavazza, A., S. Mongey, and G. L. Violante (2018): "Aggregate Recruiting Intensity," *American Economic Review*, 108(8), 2088–2127.

Gertler, M., C. Huckfeldt, and A. Trigari (2019): "Unemployment Fluctuations, Match Quality, and the Wage Cyclicality of New Hires," Review of Economic Studies, forthcoming.

Gertler, M., L. Sala, and A. Trigari (2008): "An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining," *Journal of Money, Credit and Banking*, 40(8), 1713–1764.

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.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018): "Cyclical Job Ladders by Firm Size and Firm Wage," *American Economic Journal: Macroeconomics*, 10(2), 52–85.

Haltiwanger, J. C., H. R. Hyatt, E. McEntarfer, and M. Staiger (2021): "Cyclical Worker Flows: Cleansing vs. Sullying," NBER Working Papers 28802, National Bureau of Economic Research, Inc.

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022): "The Slope of the Phillips Curve: Evidence from U.S. States," Quarterly Journal of Economics, forthcoming.

Holden, T. (2011): "Products, patents and productivity persistence: A DSGE model of endogenous growth," Dynare Working Papers 4, CEPREMAP.

Holden, T., and M. Paetz (2012): "Efficient Simulation of DSGE Models with Inequality Constraints," Quantitative Macroeconomics Working Papers 21207b, Hamburg University, Department of Economics.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2010): "Investment shocks and business cycles," *Journal of Monetary Economics*, 57(2), 132–145.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27(4), 1241–1277.

Krause, M. U., D. Lopez-Salido, and T. A. Lubik (2008): "Inflation dynamics with search frictions: A structural econometric analysis," *Journal of Monetary Economics*, 55(5), 892–916.

Levin, A., J. D. Lopez-Salido, and T. Yun (2007): "Strategic Complementarities and Optimal Monetary Policy," CEPR Discussion Papers 6423, C.E.P.R. Discussion Papers.

Lindé, J., F. Smets, and R. Wouters (2016): "Chapter 28 - Challenges for Central Banks Macro Models," vol. 2 of *Handbook of Macroeconomics*, pp. 2185–2262. Elsevier.

Lindé, J., and M. Trabandt (2018): "Resolving the Missing Deflation Puzzle," mimeo, Freie Universitat Berlin.

McLeay, M., and S. Tenreyro (2019): "Optimal Inflation and the Identification of the Phillips Curve," in *NBER Macroeconomics Annual 2019, volume 34*. National Bureau of Economic Research, Inc.

Moscarini, G., and F. Postel-Vinay (2016): "Did the Job Ladder Fail after the Great Recession?," *Journal of Labor Economics*, 34(S1), 55–93.

——— (2018): "On the Job Search and Business Cycles," Mimeo Yale U. and U. College London.

——— (2019): "The Job Ladder: Inflation vs Reallocation," Mimeo Yale U. and U. College London.

Phillips, W. (1958): "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861 - 1957," *Economica*, 25(100), 293–299.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Ravenna, F., and C. E. Walsh (2008): "Vacancies, unemployment, and the Phillips curve," *European Economic Review*, 52(8), 1494–1521.

Shimer, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *The American Economic Review*, 95(1), 25–49.

Silva, J. I., and M. Toledo (2009): "Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment," *Macroeconomic Dynamics*, 13(S1), 76–96.

Sims, C., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–68.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97(3), 586–606.

Stock, J. H., and M. W. Watson (2007): "Why Has U.S. Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 39(s1), 3–33.

Stock, J. H., and M. W. Watson (2008): "Phillips Curve Inflation Forecasts," Working Paper 14322, National Bureau of Economic Research.

——— (2019): "Slack and Cyclically Sensitive Inflation," Working Paper 25987, National Bureau of Economic Research.

# Appendices (Not For Publication)

In Appendix A, we show the EE/UE in the data. We summarize how to construct the measure of marginal costs in a standard New Keynesian model in Appendix B. Different calibrations and specifications for the Phillips curve studied in Section 2 of the main text are introduced, and their ability to account for the missing inflation after the Great Recession is evaluated in Appendix C, which focuses on Phillips curves with a backward-looking component. In Appendix D we describe how the data set to conduct the VAR analysis in Section 2 of the main text is constructed. In Appendix E, we show that state-of-the-art dynamic general equilibrium models have hard time explaining the missing inflation. We show how to work out equations (13) and (14) in the main text, which provide an analytical characterization of the surpluses in the model, in Appendix F. Appendix G illustrates analytically the link between inflation and the labor market states of unemployment, on-the-job search and bad matches. In Appendix H, we show the robustness of our main results by varying two parameters that are hard to calibrate: the probability of meeting a worker that is a bad match for the firm ($\xi_b$) and the probability that workers switch jobs if they receive an outside offer that makes them indifferent ($\nu$). We also show how the results change when varying the persistence of the on-the-job search rate ($\rho_s$). In Appendix I, we show how we solve the model with an occasionally binding zero lower bound for the nominal interest rate. In Appendix J we show how preference shocks propagate. Appendix K compares the performance of our indicator of interfirm wage competition against traditional measures of slack. Appendix L presents robustness exercises where the model is taken to alternative sets of labor market variables. Appendix M presents more in details the empirical strategy used to take the model to the period of the pandemic. We provide more specifics about the model in which agents optimally decide whether to search on the job and its predictions in Appendices N, O and P. Finally, Appendix Q discusses the limitations of our empirical approach in the identification of the rate of on-the-job search.

# A Acceptance Rate

Figure 11 shows the ratio of the employment-to-employment flow rate, corrected as suggested by Fujita, Moscarini, and Postel-Vinay (2019), to the unemployment-to-employment flow rate.[33] This plot shows that the EE/UE rate rapidly rose during the Great Recession. However, this rate steadily decreased during the recovery and eventually moved below its pre-Great Recession average computed over the period from February 1996 through December 2007, which is denoted

[33]The correction proposed by Fujita, Moscarini, and Postel-Vinay (2019) ends up revising the employment-to-employment rate upward in recent years, causing the fall of this ratio to be less rapid and dramatic during the post-Great Recession recovery than one would obtain by using the uncorrected CPS series for the EE flow rate.

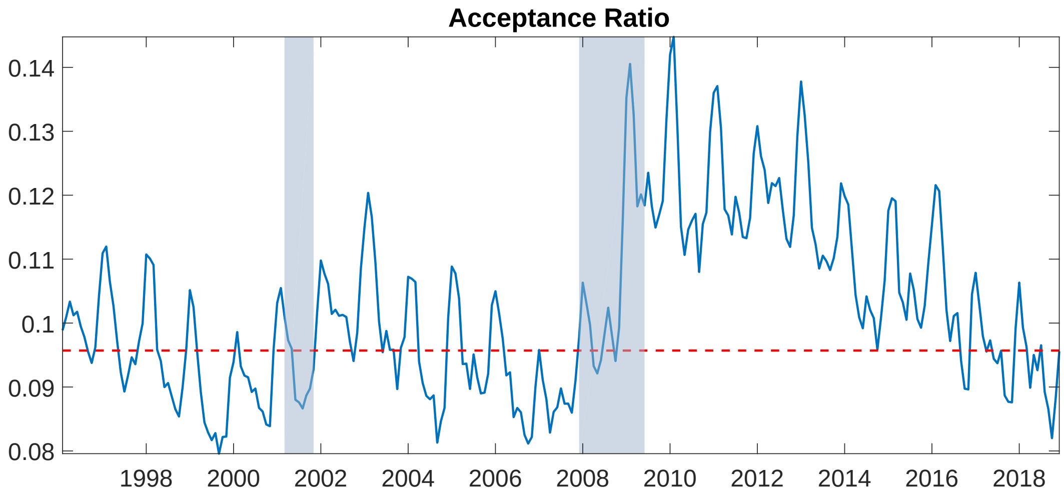


Figure 11: Acceptance Ratio. The ratio of the employment-to-employment flow rate to the unemployment-to-employment flow rate. Both rates are computed by taking the three-month moving average of the Current Population Survey (CPS) flow data. The red dashed line denotes the mean of the ratio computed from February 1996 through December 2007. The employment-to-employment rate is corrected as proposed by Fujita, Moscarini, and Postel-Vinay (2019).

by the red dashed line.[34]

Moscarini and Postel-Vinay (2019) interpret this rate as an acceptance rate. In their model, because the fraction of accepted offers is higher when more workers are employed in low-productivity jobs, this rate is a proxy for the degree of labor misallocation and is inversely related to inflation. When this rate is low, few offers are accepted on average as labor is perfectly allocated and, as result, marginal costs and inflation are high in their model. In our model, a low EE/UE rate may be due to either a high degree of misallocation or a low share of workers searching on the job. Therefore, this rate is not always a good predictor of labor misallocation and inflation in our model. A better predictor is the empirical measure of labor market slack, which is based on the intensity of interfirm wage competition, introduced in Section 3.6.

# B Computation of Real Marginal Costs in a Standard New Keynesian Model with Search and Matching

We follow the work by Krause, Lopez-Salido, and Lubik (2008), who study the behavior of real marginal costs in a simple New Keynesian model with search and matching frictions in the labor market. Equation (32) from Krause, Lopez-Salido, and Lubik (2008, p. 898) defines the

[34]The CPS data start in February 1996.

real marginal cost as:

$$mc_t = \frac{W_t}{\alpha\left(\frac{y_t}{n_t}\right)} + \frac{c'\left(v_t\right)/q\left(\theta_t\right) - \left(1-\rho\right)E_t\beta_{t+1}c'\left(v_{t+1}\right)/q\left(\theta_{t+1}\right)}{\alpha\left(\frac{y_t}{n_t}\right)}, \tag{26}$$

where $W_t$ denotes the real hourly wage, $y_t/n_t$ is the average product of labor, $c'\left(v_t\right)$ is the derivative of the vacancy cost function with respect to vacancies, $q\left(\theta_t\right)$ is the vacancy filling rate, $\beta_{t+1}$ is the discount factor, and $\alpha$ is the elasticity of output to employment in the production function. The first component on the right-hand side of equation (26) is the unit labor cost, i.e., the ratio of the labor cost and the marginal product of labor. The second component stems from the existence of search and matching frictions and can be interpreted as cost savings from not having to hire in the following period.

Let $s_t \equiv W_t/\alpha\left(\frac{y_t}{n_t}\right)$ denote the unit labor cost, which equals the labor share of income divided by the elasticity of output to employment. Krause, Lopez-Salido, and Lubik (2008) show that linearizing equation (26) and rearranging leads to the following expression:

$$\widehat{mc_t} = \hat{s}_t + \frac{1-\phi}{1-\tilde{\beta}}\left[\frac{\xi}{1-\xi}\left(\hat{h}_t - \tilde{\beta}E_t\hat{h}_{t+1}\right) + \left(\varepsilon_c - 1\right)\left(\hat{v}_t - \tilde{\beta}E_t\hat{v}_{t+1}\right) - \tilde{\beta}E_t\hat{\beta}_{t+1} - \left(1-\tilde{\beta}\right)\hat{w}_t\right], \tag{27}$$

where a hat variable is used to denote log deviations from the steady-state, $h_t$ denotes the job finding rate, $\tilde{\beta}$ is a discount factor adjusted for the rate of job separation, $\varepsilon_c$ is the elasticity of vacancy costs to vacancies, $\xi$ is the elasticity of the matching function with respect to unemployment, and $\phi = s/mc$ is the share of unit labor cost over total marginal costs. We follow the calibration in Krause, Lopez-Salido, and Lubik (2008) and assume that $\xi = 0.5$, $1-\phi = 0.05$, and $\tilde{\beta} = 0.943$. In line with the model specified in Section 3, we assume a linear vacancy cost function, which implies $\varepsilon_c = 1$, and log utility in consumption.

# C Traditional Measure of Slack: Robustness

The most popular Phillips curve used in empirical studies features a backward-looking term:

$$\pi_t = \iota\pi_{t-1} + \kappa\varphi_t + E\pi_{t+1}, \tag{28}$$

where the parameter $\iota$ reflects the degree of price indexation. We redo the same VAR-based exercise as the one in Section 2 to evaluate the robustness of these results to the introduction of price indexation and of a flatter Phillips curve. We kick off by setting the degree of price indexation $\iota_p$ to 0.9—an upper bound for plausible degrees of inertia. We consider two cases. In the first case, we assume that the slope of the Phillips curve is $\kappa = 0.005$, as in the baseline case analyzed in the main text. In the second case, we consider a very flat Phillips curve with a

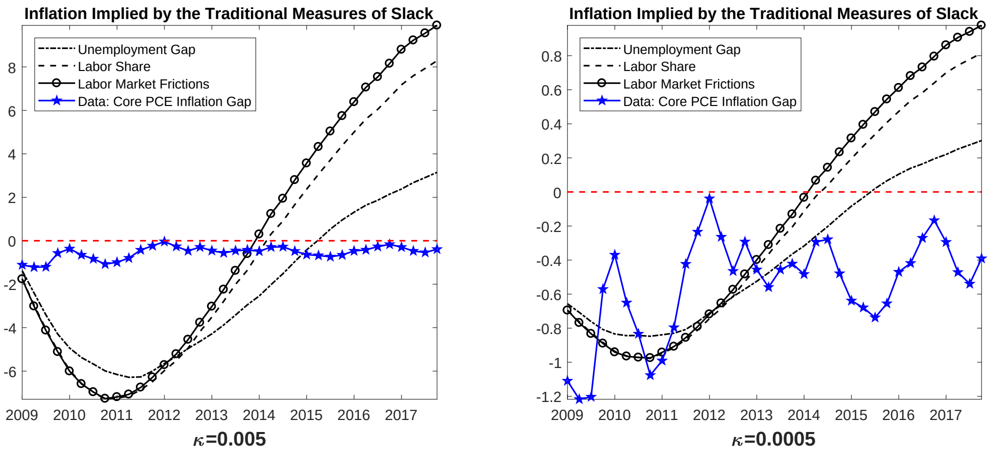


Figure 12: PCE refers to the Price Index for Personal Consumption Expenditures. The figure shows the inflation dynamics from 2009Q1 through 2017Q3 using three traditional theories of inflation using a slope of the inertial Phillips curve of 0.005 (left panel) and 0.0005 (right panel) and the core PCE inflation gap. The assumed degree of inertia is equal to 0.90.

parameter $\kappa = 0.0005$. While the first case allows us to evaluate how much adding a backward-looking component alters the result shown in the main text (Figure 2), the second case is useful to illustrate that even a very flat Phillips curve with a lot of price indexation cannot solve the missing inflation puzzle.

The results are shown in Figure 12. In the left panel we show the first case, which confirms that adding price indexation just makes the drop in inflation in 2009 more pronounced and delayed. According to the traditional measures of slack, inflation should have picked up by 2014; only the unemployment gap predicts inflation staying below its long-run level until the end of 2015. None of the measures explains why inflation has not risen after nine years of economic expansion, even after introducing a very large degree of price indexation.

In the right panel of Figure 12 we show the second case, which combines a large degree of price indexation with an extremely flat slope of the Phillips curve. Comparing the plots of the left and right panels of Figure 12 reveals that a reduction of the slope in the presence of high indexation decreases the predicted fall in inflation at the beginning of the sample and contains the predicted rise in inflation at the end of the sample. Therefore, we conclude that a flatter Phillips curve in and of itself does not solve the puzzle of the persistently low inflation observed in the past decade. This is the case even if one endows the Phillips curve with a very large price inertia.

# D Construction of the Time Series and Their Sources

The time series used for the VAR analysis have been constructed from the following data downloaded from the St. Louis Fed's database called Federal Reserve Economic Data (FRED). The labor share of income is computed as the ratio of total compensation in the nonfarm business sector divided by nominal nonfarm GDP. In turn, total compensation is computed as the product of compensation per hour (COMPNFB) times total hours (HOANBS), and nominal GDP is the product of real output (OUTNFB) times the appropriate deflator (IPDNBS). All series are quarterly and seasonally adjusted. We compute the deviations of the labor share from its trend by computing log deviations from an eight-year moving average.

We follow Shimer (2005) and compute the job finding rate as $\phi_t = 1 - \left(u_{t+1} - u^s_{t+1}\right)/u_t$, where $u^s_{t+1}$ denotes the number of workers employed for less than five weeks in month $t+1$ (UEMPLT5). The total number of workers unemployed in each month is computed as the sum of the number of civilians unemployed less than five weeks (UEMPLT5), for 5 to 14 weeks (UEMP5TO14), 15 to 26 weeks (UEMP15T26), and 27 weeks and over (UEMP27OV). The primary data are constructed by the U.S. Bureau of Labor Statistics from the CPS and seasonally adjusted. To obtain quarterly percentage point deviations of the job finding rate from its trend, we average monthly data over each quarter and then subtract the actual job finding rate from its eight-year moving average.

We also use data on real gross domestic product (GDPC1), real gross private domestic investment (GDPIC1), and real personal consumption expenditures (PCECC96). All data are quarterly and seasonally adjusted. When computing percentage deviations of each of these times series from its trend, we first remove a quadratic trend from the variable in logs and then take the difference from its eight-year moving averages. To compute percentage deviations of real wages from the trend we first remove a linear trend to the log of compensation per hour (COMPNFB) and then take the difference with respect to its eight-year moving average.

We measure aggregate price inflation by taking log differences on the previous quarter of the seasonally adjusted Consumer Price Index for All Urban Consumers (CPIAUCSL). We also use quarterly data on the effective federal funds rate (FFR) and on the short-term natural rate of unemployment (NROUST). We compute percentage point deviations of inflation, the federal funds rate, and the natural rate of unemployment from trend as the difference from each series' eight-year moving average.

# E A State-of-the-Art Dynamic General Equilibrium Model (Smets and Wouters 2007)

In this appendix, we evaluate the ability of a leading empirical general equilibrium model to reconcile labor market and inflation dynamics in the post-Great Recession recovery. We use the popular model introduced by Smets and Wouters (2007) to perform this exercise. This is a model with many real and nominal frictions and a large array of shocks and is well known to fit the U.S. macro series well. Smets and Wouters conduct a Bayesian estimation of the parameters of their model using seven observables: consumption growth, investment growth,

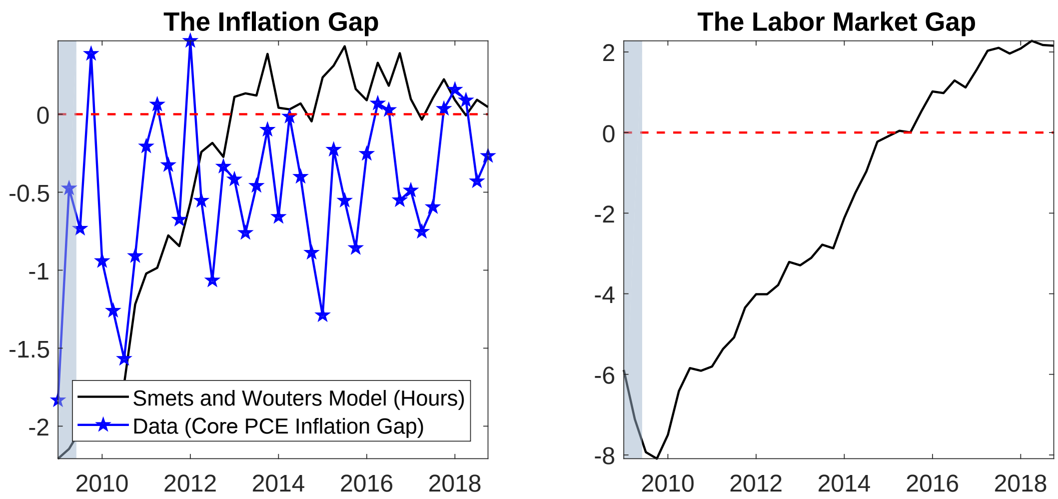


Figure 13: Left panel: PCE refers to the Price Index for Personal Consumption Expenditures. The black solid line: Inflation predicted by the model of Smets and Wouters (2007) conditional on observing the series of hours plotted on the right panel. The blue starred line: the difference between the annualized quarter-to-quarter core PCE inflation rate and the ten-year-ahead core PCE inflation expectations based on the Survey of Professional Forecasters. Right panel: Hours worked detrended using their eight-year moving average.

GDP growth, hours (detrended for the labor force participation), inflation, real wage, and the federal funds rate. Their sample period goes from 1966Q1 through 2004Q4. We extend their data set trough 2018Q4 and detrend the series of hours using a eight-year moving average. We make the latter change because the series of hours exhibited a significant downward shift since the onset of the Great Recession and has never attained its pre-recession level again.

We use the extended data set to estimate the model. Then the same data set is used to filter the state variables of the estimated model from the first quarter of 1966 through the fourth quarter of 2008. For the subsequent periods (2009Q1–2018Q4), we filter the state variables of the estimated model using only the series of hours in order to obtain inflation predictions conditional on labor market data only. Recall that the emphasis of this paper is on the apparently waning link between the labor market and inflation. The black solid line in the right panel of Figure 13 shows the series of hours detrended using a eight-year moving average, which we use to simulate the Smets and Wouters model.

Based on the series of hours, the Smets and Wouters' model predicts that inflation is above target already in 2012. See the black solid line in the left panel of Figure 13. The plot also reports the inflation gap in the data (blue starred line), which is computed by taking the difference between the annualized quarter-to-quarter core PCE inflation rate and the ten-year-ahead core PCE inflation expectations based on the *Survey of Professional Forecasters*. The inflation gap in the data remains persistently below zero, whereas the Smets and Wouters' model predicts that inflation moves above its long-run level as early as in 2012. Indeed, the right panel of Figure 13 shows that the series of hours implied that the labor market became tight (positive labor market gap) in 2015.

# F Job Values and Sequential Auctions

In this section we derive the expressions for the surplus function $S_t(y)$ in equation (13), following the approach in Moscarini and Postel-Vinay (2019). We start by characterizing the value functions for the states of employment and unemployment. The value of unemployment to a worker $j$ measured after worker reallocation has taken place and expressed in utility units is determined as follows::

$$\lambda_t V_{u,t}^j = b + \beta E_t \phi(\theta_{t+1}) \lambda_{t+1} \left[ V_{e,t+1}^j \left( w_{t+1}(j), y_{t+1}(j) \mid e_{t+1}^0 = 0 \right) \right] + \beta E_t (1 - \phi(\theta_{t+1})) \lambda_{t+1} V_{u,t+1}^j, \tag{29}$$

where we let the indicator function $e_{t+1}^0 = \{0, 1\}$ denote the state of employment at the beginning of period $t+1$, before reallocation takes place.

The value to a worker $j$ of being employed at the production stage of period $t$ in a job of productivity $y_t$ at wage $w_t$ after reallocation has taken place, but before the realization of the current-period separation shock, is determined as follows:

$$\begin{aligned} \lambda_t V_{e,t}^j (w_t(j), y_t(j)) &= \lambda_t \frac{w_t(j)}{P_t} + \beta E_t \lambda_{t+1} \left\{ \delta \left[ 1 - \phi(\theta_{t+1}) \right] V_{u,t+1}^j \right. \\ &+ \delta \phi(\theta_{t+1}) V_{e,t+1}^j \left( w_{t+1}(j), y_{t+1}(j) \mid e_{t+1}^0 = 0 \right) \\ &\left. + (1 - \delta) V_{e,t+1}^j \left( w_{t+1}(j), y_{t+1}(j) \mid w_t(j), y_t(j), e_{t+1}^0 = 1 \right) \right\}. \end{aligned} \tag{30}$$

The above expression implies that the worker receives a wage $\frac{w_t(j)}{P_t}$ in exchange for her labor services, plus a continuation value, which depends on whether the worker separates or not at the end of the period. If separation occurs at rate $\delta$, the worker will still be in the state of unemployment by the end of period $t+1$ if no job is found, which occurs with probability $1 - \phi(\theta_{t+1})$. In this case the worker receives the expected present value $E_t V_{u,t+1}^j$. If instead the newly separated worker finds a job in period $t+1$ with probability $\phi(\theta_{t+1})$, she gets the payoff of being in a match of productivity $y_{t+1}(j)$, paying the wage $w_{t+1}(j)$, which is conditional on the worker having separated at the end of time $t$ and therefore being unemployed at the beginning of $t+1$. The expected present discounted value of a such job, expressed in units of the numeraire good is denoted by $E_t V_{e,t+1} \left[ w_{t+1}(j), y_{t+1}(j) \mid e_{t+1}^0 = 0 \right]$. With probability $1 - \delta$ instead, the worker does not separate at the end of time $t$, receiving $E_t V_{e,t+1} \left[ w_{t+1}(j), y_{t+1}(j) \mid w_t(j), y_t(j), e_{t+1}^0 = 1 \right]$ at the end of the next period. This expression captures the value of being employed at the end of time $t+1$ in a match with productivity $y_{t+1}$ at the wage $w_{t+1}$, conditional on having been employed in a match with productivity $y_t(j)$ and wage $w_t(j)$ in the previous period and not having separated between periods, i.e., being in employment at the beginning of period $t+1$. Note that this expected value includes the possibility of a job-to-job transition in period $t+1$.

We assume that firms have all the bargaining power, and hence, the unemployed workers who take up a new offer are indifferent between being employed or unemployed, i.e.,

$$\lambda_t V_{e,t}\left(w_t(j), y_t(j) \mid e_t^0 = 0\right) = b + \beta E_t \lambda_{t+1} V_{u,t+1}^j \tag{31}$$

independently of $y_t(j)$. It follows that

$$V_{u,t}^j = \frac{b}{\lambda_t} + \beta E_t \frac{\lambda_{t+1}}{\lambda_t} V_{u,t+1}^j = V_{u,t}. \tag{32}$$

Let $V_{e,t}^*(y)$ denote the value to the worker of being employed under full extraction of a firm's willingness to pay at the end of time $t$. In this case a worker of productivity $y$ receives the maximum value that the firm is willing to promise in period $t$, including the payment of the current-period wage. Let $\{w_s^*(y)\}_{s=t}^{\infty}$ denote the state-contingent contract that delivers $V_{e,t}^*(y) \equiv V_{e,t}(w_t^*, y)$. By promising to pay the contract $\{w_s^*(y)\}_{s=t}^{\infty}$, the firm breaks even in expectation, that is, the expected present value of future profits is zero.

Now consider a firm that is currently employing a worker with productivity $y$ under any promised contract $\{w_s(y)\}_{s=t}^{\infty}$. Assume that the worker is poached by a firm with match productivity $y'$. The outcome of the auction must be one of the following three:

1. $V_{e,t}^*(y') < V_{e,t}(w_t, y)$; in this case the willingness to pay of the poaching firm is less than the value of the contract that the worker is currently receiving. As a result, the incumbent firm retains the worker with the same wage contract with value $V_{e,t}(w_t, y)$.

2. $V_{e,t}(w_t, y) \leq V_{e,t}^*(y') < V_{e.t}^*(y)$; in this case the willingness to pay of the poaching firm is greater or equal to the value of the contract the worker is receiving in his current job, but lower than the willingness to pay of the incumbent firm. The two firms engage in Bertrand competition, and as a result, the incumbent firm retains the worker offering the new contract $V_{e,t}^*(y')$.

3. $V_{e,t}^*(y) \leq V_{e,t}^*(y')$; in this case the poaching firm has a willingness to pay that is no less than the incumbent's. If this condition holds with strict inequality, the current match is terminated and the worker is poached at the maximum value of the contract that the incumbent is willing to pay. If instead the worker is poached by a firm with equal productivity, it is assumed that job switching takes place with probability $\nu$. In either case, the continuation value of the contract obtained by the worker is $V_{e,t}^*(y)$.

The bargaining protocol above, together with the assumption that entrant firms make zero profits in expectations, yields the free entry-condition, i.e., equation (12) in the main text, which we display again below for convenience:

$$
\begin{aligned}
c^f + \frac{c}{\varpi_t} &= \frac{u_{0,t}}{u_{0,t} + s_t\left(1 - u_{0,t}\right)} \left\{\xi_b S_t\left(y_b\right) + \xi_g S_t\left(y_g\right)\right\} \\
&\quad + \frac{s_t\left(1 - u_{0,t}\right)}{u_{0,t} + s_t\left(1 - u_{0,t}\right)} \left\{\xi_g \frac{l_{b,t}^0}{1 - u_{0,t}} \left[S_t\left(y_g\right) - S_t\left(y_b\right)\right]\right\}.
\end{aligned} \tag{33}
$$

Substituting out for the surplus functions in the above equations requires some steps. Start by considering the case of a firm that has promised to pay the contract $\{w_s^*(y)\}_{s=t}^{\infty}$, which implies that the firm breaks even in expectation and is not able to promise higher wage payments in case it enters an auction with a poaching firm. In this case, if no outside offers arrive the worker receives a continuation value of $V_{e,t}^*(y)$ from the incumbent firm. Otherwise the worker is poached and, in accordance with point (3) above, receives from the new firm a contract that is also worth $V_{e,t}(w', y') = V_{e,t}^*(y)$. So either way, the worker receives a contract of value $V_{e,t}^*(y)$. The value to a worker of being employed under the contract $\{w_s^*(y)\}_{s=t}^{\infty}$ can therefore be written as:

$$
V_{e,t}^*(y) = \varphi_t y + \beta E_t \frac{\lambda_{t+1}}{\lambda_t} \left[\delta V_{u,t} + (1-\delta) V_{e,t+1}^*(y)\right], \tag{34}
$$

where $\varphi_t y$ is the marginal revenue product of selling $y$ units of the service to the price setters. Subtracting (32) from the above equation yields:

$$
V_{e,t}^*(y) - V_{u,t} = \varphi_t y - \frac{b}{\lambda_t} + (1-\delta)\beta E_t \frac{\lambda_{t+1}}{\lambda_t} \left[V_{e,t+1}^*(y) - V_{u,t+1}\right]. \tag{35}
$$

Notice that the value to the worker of extracting all the rents associated with a type-$y$ match, $V_{e,t}^*(y) - V_{u,t}$, is in fact simply the surplus $S_t(y)$. Iterating forward on the above expression, we can define the surplus of a match with productivity $y$ as:

$$
S_t(y) = E_t\left[\sum_{\tau=0}^{\infty} (1-\delta)^{\tau} \left(\frac{\lambda_{t+\tau}}{\lambda_t} \varphi_{t+\tau} y - \frac{b}{\lambda_t}\right)\right]. \tag{36}
$$

Notice that the surplus function above is affine increasing in $y$, which implies that firms with higher productivity win the auction and, therefore, workers cannot move to jobs with lower productivity. For convenience, we can rearrange the above expression as

$$
S_t(y) = y\mathcal{W}_t - \frac{b\lambda_t^{-1}}{1 - \beta(1-\delta)}, \tag{37}
$$

where

$$
\mathcal{W}_t = \varphi_t + \beta(1-\delta) E_t \frac{\lambda_{t+1}}{\lambda_t} \mathcal{W}_{t+1}. \tag{38}
$$

Seen from the point of view of a service sector firm, $\mathcal{W}_t$ can be interpreted as the expected present discounted value of the entire stream of current and future real marginal revenues derived from selling one unit of the service until separation. From the point of view of a price setting firm, which purchases labor services, $\mathcal{W}_t$ can be interpreted as the expected present discounted value of the cost of purchasing one unit of the labor service by a firm until separation.

Using equation (37) we can now substitute for the surplus functions and rearrange to rewrite the free-entry condition (12) as:

$$
\begin{aligned}
c^f + \frac{c}{\varpi_t} \quad = \quad & \frac{u_{0,t}}{u_{0,t} + s_t\left(1 - u_{0,t}\right)}\left[\mathcal{W}_t\left(\xi_b y_b + \xi_g y_g\right) - \frac{b\lambda_t^{-1}}{1 - \beta\left(1 - \delta\right)}\right] \\
& + \frac{s_t}{u_{0,t} + s_t\left(1 - u_{0,t}\right)}\xi_g l_{b,t}^0 \mathcal{W}_t\left(y_g - y_b\right).
\end{aligned} \tag{39}
$$

# G An Analytical Derivation of the Drivers of Inflation in the Model

To understand how labor market conditions affect inflation in the model, it is useful to inspect the link between $\mathcal{W}_t$ in equation (14) of the main text,

$$
\mathcal{W}_t = \varphi_t + \beta\left(1 - \delta\right) E_t \frac{\lambda_{t+1}}{\lambda_t}\mathcal{W}_{t+1}, \tag{40}
$$

and inflation, whose dynamics are governed by the standard New Keynesian Phillips curve in the model,

$$
\hat{\pi}_t = \kappa\hat{\varphi}_t + \beta E\hat{\pi}_{t+1}, \tag{41}
$$

where we use the hat symbol $\widehat{}$ over a variable to denote deviations from the steady state. Assume, for simplicity, that $\beta \simeq 1$ and the separation rate is negligibly small, i.e. $\delta \simeq 0$. Then loglinearizing equation (40) yields:

$$
\widehat{\Omega}_t = \kappa\hat{\varphi}_t + \beta E\widehat{\Omega}_{t+1}, \tag{42}
$$

where

$$
\widehat{\Omega}_t \equiv \frac{\mathcal{W}}{\varphi}\kappa\left[\widehat{\mathcal{W}}_t + \hat{\lambda}_t\right] \tag{43}
$$

and both $\mathcal{W}$ and $\varphi$ denote steady-state values. Comparing equations (42) and (41) implies that

$$
\hat{\pi}_t = \widehat{\Omega}_t. \tag{44}
$$

We can therefore understand how price inflation relates to labor market conditions by inspecting the mapping between $\widehat{\Omega}_t$ and the state variables that characterize the labor market. This can be done by replacing the surplus function (equation 13) and the definition of vacancy filling rate $\varpi_t = \frac{\theta_t}{\phi(\theta_t)}$ into the free entry condition (equation 12), log linearizing, and solving for $\widehat{\mathcal{W}}_t$.

So we first express the free entry condition as:

$$\begin{aligned} c^f + c\theta_t &= \frac{u_{0,t}\phi(\theta_t)}{u_{0,t} + s(1-u_{0,t})}\left[\mathcal{W}_t\left(\xi_b y_b + \xi_g y_g\right) - \frac{b}{\lambda_t}\frac{1}{1-\beta(1-\delta)}\right] \\ &\quad + \frac{s_t\phi(\theta_t)}{u_{0,t} + s_t(1-u_{0,t})}\mathcal{W}_t\xi_g l^0_{b,t}\left[y_g - y_b\right]. \end{aligned} \tag{45}$$

Then we loglinearize the above equation to get:

$$\begin{aligned} &\left\{c\theta - \frac{u_0}{u_0 + s(1-u_0)}\phi'(\theta)\left[\mathcal{W}\left(\xi_b y_b + \xi_g y_g\right) - \frac{b}{\lambda}\frac{1}{1-\beta(1-\delta)}\right]\theta \right. \\ &\left. - \frac{s}{u_0 + s(1-u_0)}\phi'(\theta)\xi_g\left[y_g - y_b\right]l^0_b\mathcal{W}\theta\right\}\cdot\hat{\theta}_t \\ &- \left\{\frac{u_0}{u_0 + s(1-u_0)}\phi(\theta)\left[\frac{b}{\lambda}\frac{1}{1-\beta(1-\delta)}\right]\right\}\cdot\hat{\lambda}_t \\ &- \left\{\frac{u_0}{u_0 + s(1-u_0)}\phi(\theta)\left[\mathcal{W}\left(\xi_b y_b + \xi_g y_g\right)\right] + \frac{s\phi(\theta)}{u_0 + s(1-u_0)}\xi_g\left[y_g - y_b\right]l^0_b\mathcal{W}\right\}\widehat{\mathcal{W}_t} \\ &- \left\{\frac{s\phi(\theta)}{u_0 + s(1-u_0)}\xi_g\left[y_g - y_b\right]l^0_b\mathcal{W}\right\}\cdot\hat{l}^0_{b,t} \\ &- \left\{\frac{u_0 s}{\left[u_0 + s(1-u_0)\right]^2}\phi(\theta)\left[\xi_g S_g\left[1 - \frac{l^0_b}{1-u_0}(1-s)(1-u_0)\right]\right]\right. \\ &\left. + \xi_b S_b + \frac{l^0_b}{1-u_0}(1-s)(1-u_0)S_b\right\}\cdot\hat{u}_{0,t} \\ &- \left\{\frac{su_0(1-u_0)}{\left[u_0 + s(1-u_0)\right]^2}\phi(\theta)\left[\xi_b S_b + \xi_g S_g\left(1 - \frac{l^0_b}{1-u_0}\right) + \xi_g\frac{l^0_b}{1-u_0}S_b\right]\right\}\hat{s}_t = 0, \end{aligned}$$

and rewrite the above expression as:

$$\widehat{\mathcal{W}_t} = a_\theta\hat{\theta}_t - a_\lambda\hat{\lambda}_t - a_b\hat{l}^0_{b,t} - a_u\hat{u}_{0,t} + a_s\hat{s}_t. \tag{46}$$

where the coefficient $a_i$ for $i = \{\theta, \lambda, b, u, s\}$ are the ratios of the expressions in the curly brackets that multiply the variable $\hat{\theta}_t, \hat{\lambda}_t, \hat{l}^0_{b,t}, \hat{u}_{0,t}, \hat{s}_t$, respectively, and divided by the positive bracketed expression that premultiplies $\widehat{\mathcal{W}_t}$. Note that $a_b > 0$, $a_u > 0$ and $a_s > 0$, which implies that $\frac{\partial\widehat{\mathcal{W}_t}}{\hat{l}^0_{b,t}} < 0$, $\frac{\partial\widehat{\mathcal{W}_t}}{\hat{u}_{0,t}} < 0$ and $\frac{\partial\widehat{\mathcal{W}_t}}{\hat{s}_t} > 0$.

Substituting equation (46) into (43), we can express the latter as

$$\widehat{\Omega}_t = \frac{\mathcal{W}}{\varphi}\kappa\left[a_\theta\hat{\theta}_t - (a_\lambda - 1)\hat{\lambda}_t - a_b\hat{l}^0_{b,t} - a_u\hat{u}_{0,t} + a_s\hat{s}_t\right].$$

Here we note that under our calibration, the values of the coefficient $a_\theta$ is close to zero, whereas the coefficient and $a_\lambda$ is close to 1, so the effect of variations in labor market tightness and the

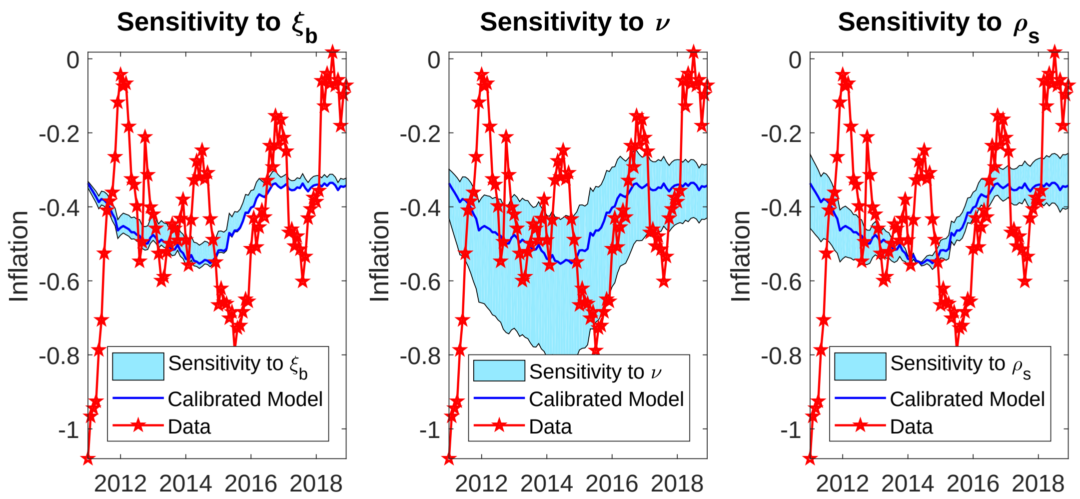


Figure 14: Robustness. Left panel: The shaded area show the sensitivity of the model's predicted year-over-year inflation rate to changes in the probability that the meeting between the worker and the firm generates a bad match ($\xi_b$). The blue solid line denotes the model's predicted year-over-year inflation rate for our baseline calibration shown in Table 1. The red starred line denotes the year-over-year inflation rate in the data (core inflation according to the Price Index for Personal Consumption Expenditures, or PCE) in deviations from the Survey of Professional Forecasters' PCE inflation expectations over the next ten years. The middle and right panels show the same plot when we perturb the probability that workers accept an offer if they are indifferent ($\nu$) and the persistence of the on-the-job search rate ($\rho_s$).

discount factor is quantitatively irrelevant in explaining changes in $\widehat{\Omega}_t$.[35] It therefore follows that

$$\widehat{\Omega}_t = \hat{\pi}_t \simeq -a_b \hat{l}^0_{b,t} - a_u \hat{u}_{0,t} + a_s \hat{s}_t, \tag{47}$$

The equation above implies that $\widehat{\Omega}_t$ and $\hat{\pi}_t$ (by virtue of equation 44) should be explained almost entirely by time variation in $\hat{l}^0_{b,t}$, $\hat{u}_{0,t}$ and $\hat{s}_t$. In Section (4.4) we have verified that this is indeed the case in our calibrated model: the simulated changes in $\hat{l}^0_{b,t}$, $\hat{u}_{0,t}$ and $\hat{s}_t$ explain 96% of the simulated variations in $\hat{\pi}_t$. It follows that in the calibrated model, the three labor market variables $\hat{l}^0_{b,t}$, $\hat{u}_{0,t}$ and $\hat{s}_t$ at time $t$, carry approximately the same informational content as the expected stream of current and future real marginal costs measured at the same point in time.

# H Robustness to Changes in Parameter Values

The shaded area in the panels of Figure 14 shows how the model's prediction of inflation changes as we vary the probability of meeting a worker that is a bad match for the firm $\xi_b$ (left), the probability that workers switch jobs if they receive an outside offer that makes them indifferent ($\nu$) (middle), or the persistence of the on-the-job search rate ($\rho_s$) (right). We consider values of the parameter $\xi_b$ ranging from 0.6 through 0.8, values of the parameter $\nu$ ranging from

[35]The coefficient $a_\theta$ is close to zero when the share of variable vacancy posting costs is small relative to the fixed cost of hiring, in line with the empirical evidence discussed in Section (4.1). In this case, changes in the cost of posting vacancies is not an important driver of the cost of labor.

0.25 through 0.75, and values of the parameter $\rho_s$ ranging from 0 through 0.97 (the highest confidence bound when the AR parameter of the series of the-on-the-job search $\tilde{s}_t$ is estimated by OLS). The blue solid line and the red starred lines denote the model's predicted inflation rate and the core PCE inflation gap for the baseline calibration reported in Table 1, respectively. These lines are the same as the ones plotted in the left panel of Figure 6.

# I Solving the Model with the ZLB Constraint

After being solved, our linearized model with the occasionally binding ZLB constraint in equation (19) can be represented in state-space form as follows:

$$s_t = \Gamma_0 s_{t-1} + \Gamma_1 \varepsilon_t^1 + \Gamma_2 \varepsilon_t^2 \tag{48}$$

where the first $k+1$ rows of $s_t$ contain the current policy rate and the expectations of the policy rate in quarter $t+1, ..., t+k$. The model's structural shocks are contained in $\varepsilon_t^2$. This vector of shocks includes the preference shock and the shocks to the on-the-job search rate. The linear system above also features a vector of dummy shocks $\varepsilon_t^1$. These shocks in $\varepsilon_t^1$ are appended to the Taylor rule so that the constrained Taylor rule in equation (19) can be written as

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_r} \left[\left(\frac{\Pi_t}{\Pi^*}\right)^{\phi_\pi} \left(\frac{Q_t}{Q^*}\right)^{\phi_y}\right]^{1-\rho_r} + \sum_{j=0}^{k} \eta_{t-j}^j, \tag{49}$$

where $\eta_t^j$ are $k+1$ monetary shocks that are known by agents at time $t$ and will hit the economy at time $t+j$. These shocks belong to the vector $\varepsilon_t^1$ in equation (48). These dummy shocks serve the sole purpose of enforcing the ZLB constraint (i.e., prevent agents from expecting negative nominal interest rates in any state of the world). Thus, the realizations of these dummy shocks will be equal to zero in every states of the world in which the current and expected nominal interest rates do not violate the ZLB constraint. It should be noted that the matrix $\Gamma_1$ is a matrix with $k+1$ columns.

As explained in the main text, the shocks are obtained by inverting the $2 \times 2$ square matrix $Z\Gamma_2$, where the matrix $Z$ is a $2 \times 2$ observation matrix such that $Y_t = Zs_t$ with the vector $Y_t$ including the observables (i.e., the unemployment rate and the EE flow rate) used in the empirical exercise whose results are described in Section 4.3 and Section 4.5. Under the assumption that the matrix $Z\Gamma_2$ is invertible (as it is in our case), this inversion allows us to retrieve the sequence of shocks $\varepsilon_t^2$ that identically explains the observed rate of unemployment and the EE rate.

We start by setting $t = 1$, which denotes the first period of our sample $Y_t$, and go through the following steps:

1. Given the realization of the two shocks $\varepsilon_t^2$ at time $t$, we set the matrix $\Psi(0) = \mathbf{0}_{k+1\times k+1}$, $\varepsilon_t^1(0) = \mathbf{0}_{k+1\times 1}$, $i = 0$, and go to Step 2.

2. Define the vector of adjustments to forward guidance shocks $\Delta\varepsilon_t^1$ that ensures the current and/or the expected path of the future interest rates will respect the ZLB as follows:

$$\Delta\varepsilon_t^1 = \left(\Gamma_1^{(0:k)}\right)^{-1}\left[-\ln R_* - \Gamma_0^{(0:k)} s_{t-1} - \Gamma_1^{(0:k)}(i)\cdot\Psi(i)\,\varepsilon_t^1(i) - \Gamma_2^{(0:k)}\varepsilon_t^2\right], \tag{50}$$

where $\Gamma_1^{(0:k)}$ denotes the square submatrix made of the first $k+1$ rows of the matrix $\Gamma_1$. With $\Delta\varepsilon_t^1$ at hand, we update $\varepsilon_t^1(i+1) = \varepsilon_t^1(i) + \Delta\varepsilon_t^1$. Note that if the ZLB constraint is not binding at time $t$, $\Delta\varepsilon_t^1 = \mathbf{0}_{k+1\times 1}$.

3. Check if the below inequality is satisfied (the ZLB is not binding),

$$\Gamma_0^{(0:k)} s_{t-1} + \Gamma_1^{(0:k)}\cdot\Psi(i)\,\varepsilon_t^1(i+1) + \Gamma_2^{(0:k)}\varepsilon_t^2 > -\ln R_*. \tag{51}$$

We adjust the diagonal matrix of zeros and ones, $\Psi(i+1)$, so that the set of horizons at which the ZLB is binding is characterized with a value equal to one in this matrix. If $\Psi(i+1)\,\varepsilon_t^1(i+1) \neq \Psi(i)\,\varepsilon_t^1(i)$, set $i = i+1$ and go to Step 2, or else the fixed point is found and we set $\varepsilon_t^1 = \Psi(i+1)\,\varepsilon_t^1(i+1)$ and go to Step 4.

4. Compute the next period's state vector as follows:

$$s_t = \Gamma_0 s_{t-1} + \Gamma_1\varepsilon_t^1(i+1) + \Gamma_2\varepsilon_t^2. \tag{52}$$

Set $t = t+1$, and go back to Step 1.

The $s_t$ coming from equation (52) is the vector containing the model-predicted values of the state variables, which is used to generate all the empirical results of the paper.

# J Propagation of Preference Shocks

Figure 15 shows the responses to a negative preference shock. As done in the main text, we report the responses of the labor market variables (unemployment, bad matches, and good matches) at the beginning of the period, and as such they do not respond on impact by construction. When the preference shock hits, households want to save more and consume less. As a result, households' demand for the differentiated goods falls, leading to a drop in the price setters' demand for the labor service and hence in its relative price $\varphi_t$. Forward-looking price setters anticipate that marginal costs will remain low and cut their price, leading the inflation

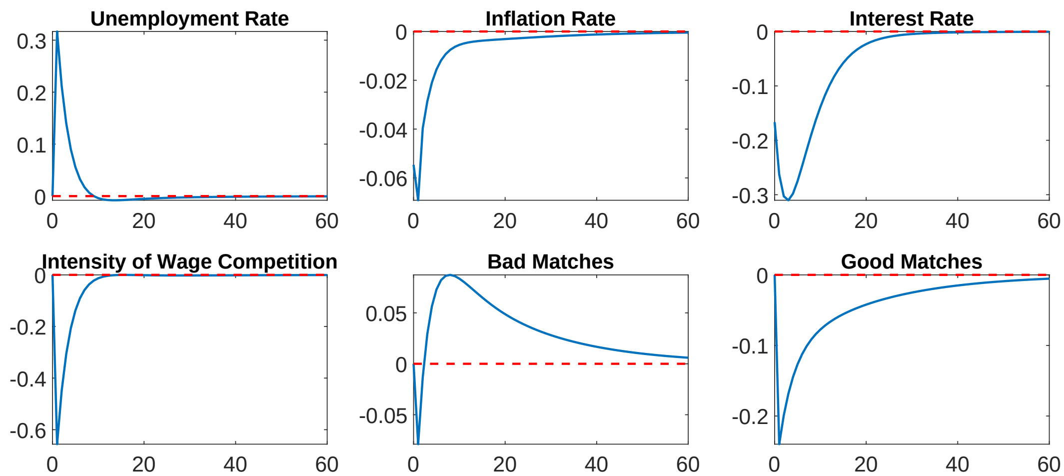


Figure 15: Impulse responses to a negative preference shock by one standard deviation. The unemployment rate, the fraction of unemployed job seekers, and the shares of good and bad matches are measured at the beginning of the period, consistently with the definition of labor market slack $\Sigma_t$ in equation (16). Units: percentage points. Inflation and interest rate are expressed in annualized rates.

rate to fall. Concurrently, the weakening of the price setters' demand for labor service reduces entry in the labor market, which in turn induces unemployment to rise over the subsequent periods. As the fraction of unemployed job seekers surges at the beginning of the second period, the expected cost of labor for an entrant service firm falls further. As a result, the intensity of wage competition, $1 - \Sigma$, where $\Sigma$ is computed based on equation (16), also falls. In turn, this leads to an even further drop in inflation in the second period.

Also note that the stock of bad matches falls initially and then rises as the entry of more labor service firms allows unemployed workers to find jobs and thus climb the ladder anew. This rise in bad matches, along with the fall in good matches, further contributes to keeping labor cheap for longer and to depressing price dynamics.

# K Empirical Performance in a Longer Sample Period

In this section, we discuss how well our measure of slack, defined in equation (16), fares at explaining inflation dynamics relative to other popular measures when the sample period is extended to include the 1990s and the 2000s. The availability of data on EE flows prevents us to extend our analysis to earlier decades.

We find that the ability of our measure of slack to explain inflation dynamics is better than the traditional ones in this longer sample period. We do so in the simplest possible way, which is to compare how our measure of slack performs relative to other traditional theory-based ones, using standard Phillips Curve regressions. That is, we estimate the equation

$$\pi_t = \beta \cdot slack_t + \varepsilon_t, \tag{53}$$

RMSE of Alternative Measures of Slack

| Measures | Sample Periods | |
|---|---|---|
| | 1992q1-2018q4 | 2011q1-2018q4 |
| Labor Share | 0.49 | 0.36 |
| Krause et al. | 0.49 | 0.40 |
| The unemployment gap | 0.49 | 0.35 |
| Detrended hours | 0.48 | 0.42 |
| Our measure of slack | 0.44 | 0.18 |

Table 2: We estimate the Phillips curve equation in the sample period and then we compute the RMSE using the estimated model for the relevant sample period.

where $\pi_t$ is the eight-quarter moving average of the quarter-over-quarter core PCE inflation rate (annualized and in percent) and in deviation from 2%, which is assumed to be the long-run value for core PCE inflation.[36] We use the moving average as we are not interested in fitting the high-frequency swings in inflation. The variable $slack_t$ represents different measures of labor market slack: our own $\Sigma_t$, based on the intensity of wage competition, and each of the measures considered in Section 2—that is, the labor share; a version of the labor share augmented to account for search and matching frictions; the unemployment gap; and detrended total hours, which is the key observable to inform the output gap in state-of-the-art DSGE models, such as those in Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2007), and Justiniano, Primiceri, and Tambalotti (2010). After having estimated the Phillips curve (53) for the period 1990Q2 through 2018Q4, we compute the root mean squared error (RMSE) of the different specifications over different subsamples.

Table 2 summarizes the Root Mean Square Errors (RMSE) of the traditional Phillips curve equation (53) estimated over different sample periods. In this longer sample period, our measure of slack attains the smallest RMSE. The particularly appalling performance of the alternative measures of slack in the last decade is the main driver of this result, justifying our emphasis on the post-Great Recession recovery in the structural analysis of this paper.

# L Robustness Exercises with Alternative Labor Market Variables in Estimation

As a first robustness check, we repeat the analysis of Section 4.5 by adding the vacancy filling rate as an observable, to our dataset that already comprises UE and EE rates.[37] In order to explain the pattern of this additional variable, we add a shock to the efficiency of the

[36] We cannot use the *Survey of Professional Forecasters*' expectations of PCE inflation over the next ten years to compute the inflation gap as we did in our empirical analysis that focused on the last decade. The reason is that this measure of long-term inflation expectations became available only since 2008.

[37] The vacancy filling rate is computed as the ratio between hires in professional and business services (JTS540099HIL) and total non-farm job openings (JTSJOL).

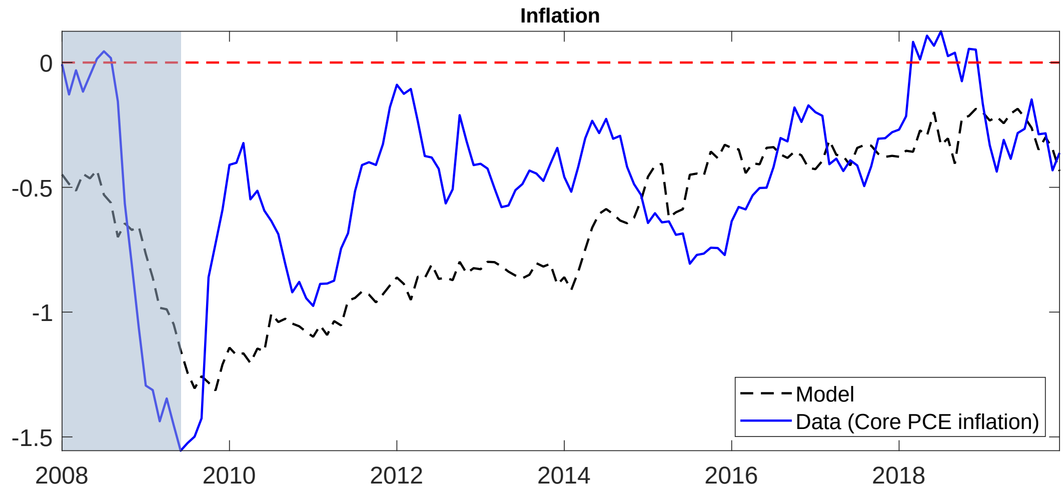


Figure 16: Model inflation vs. inflation in the data. Model inflation is obtained observing the EE flow rate from CPS and corrected as shown by Fujita, Moscarini, and Postel-Vinay (2019), the job finding rate - computed as in Hall and Schulhofer-Wohl (2018), and the job filling rate. The model is extended to feature shocks to matching efficiency.

matching function, which would capture time variation in recruiting intensity or changes to the composition of searchers over the cycle, in the spirit of Davis, Faberman, and Haltiwanger (2013) and Gavazza, Mongey, and Violante (2018), respectively. These are all factors that have been shown to contribute to improved matching of job filling and finding rates. The results, reported in Figure 16, show that even introducing such modifications to the model, the rate of inflation remains below trend by the end of the sample period.

Similarly, we provide robustness exercises where we use the U5 series for unemployment rates, instead of UE rates. This series includes among the unemployed those workers who are inactive but marginally attached to the labor force. Alternatively, we have replicated and updated the CPS-based estimates of UE rates in Hall and Schulhofer-Wohl (2018), which account for transitions from out of the labor force. The results for these two exercises, reported in Figures 17 and 18, are very similar to the baseline estimates reported in Figure 5 of the main text.

# M Estimation in the Pandemic Period

The exercise presented in this section is a variant of the exercise discussed in Section 4.3 of the main text. The equations used to extract the estimated time series for the rate of on-the-job search and the distribution of match quality are the measurement equation for the EE flow rate,

$$EE_t \equiv \frac{s_t \phi_t \left[ l^0_{b,t} \left( \xi_b \nu + \xi_g \right) + l^0_{g,t} \xi_g \nu \right]}{l^0_{b,t} + l^0_{g,t}}, \tag{54}$$

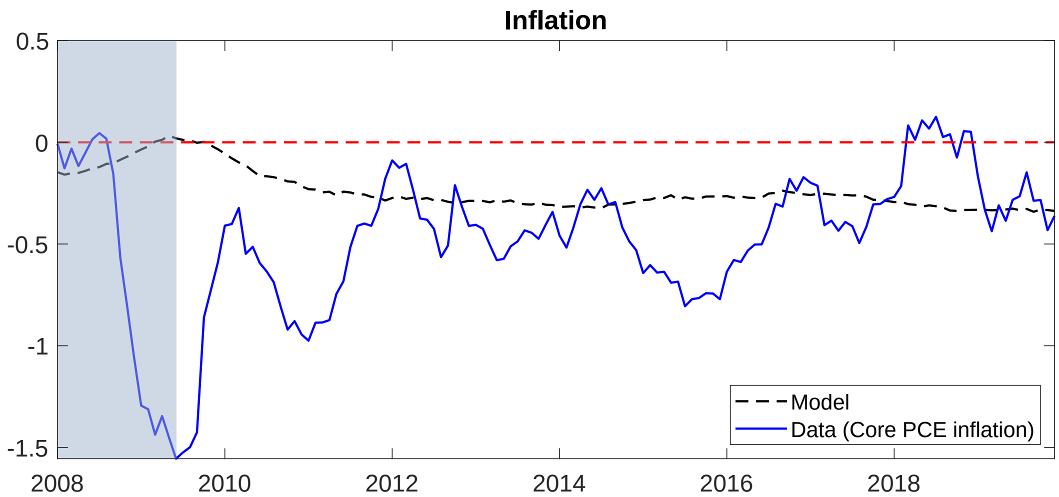


Figure 17: Model inflation vs. inflation in the data. Model inflation is obtained observing the EE flow rate and U5 unemployment rate.

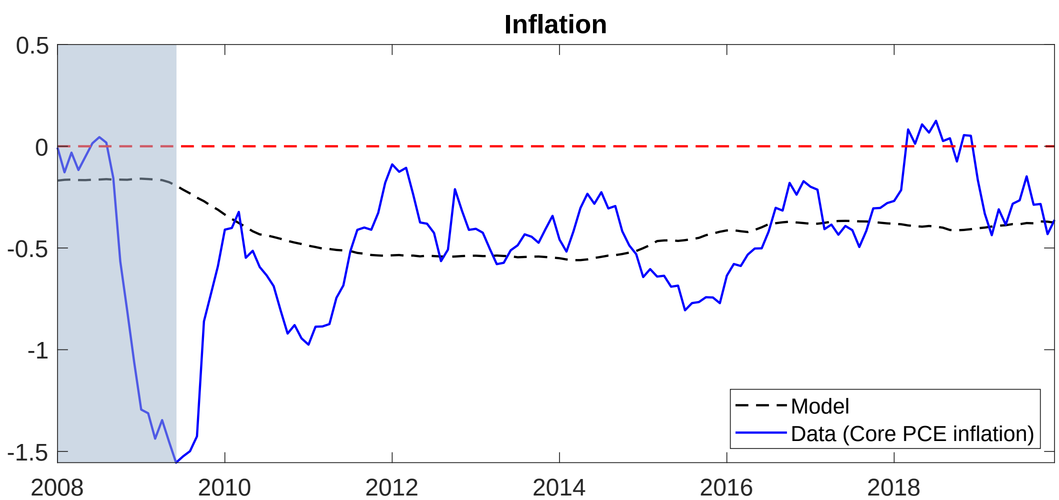


Figure 18: Model inflation vs. inflation in the data. Model inflation is obtained observing the EE flow rate and the Job finding rate estimated following Hall and Schulhofer-Wohl (2018).

the law of motion for good and bad matches,

$$l^0_{b,t+1} = (1-\delta_t)\left\{\left[1 - s_t\phi_t\xi_g\right] l^0_{b,t} + \phi_t\xi_b u_{0,t}\right\}, \tag{55}$$

$$l^0_{g,t+1} = (1-\delta_t)\left\{l^0_{g,t} + s_t\phi_t\xi_g l^0_{b,t} + \phi_t\xi_g u_{0,t}\right\}, \tag{56}$$

and the following expression for the employment-population ratio,

$$LFP^0_t\left(1-u_{0,t+1}\right) = (1-\delta_t)\left(l^0_{b,t} + l^0_{g,t} + \phi_t u_{0,t}\right). \tag{57}$$

where $LFP^0_t$ denotes the exogenous labor force participation at the beginning of the period.

The main difference, with respect to the exercise of Section 4.3 is the introduction of time-varying labor-force participation and separation rates, $\delta_t$. We derive the time series for $s_t$, $l^0_{b,t}$ and $l^0_{g,t}$ by iterating on the above equation and making $EE_t$, $u_{0,t}$, $\phi_t$ and $LFP_t$ observable in estimation. Specifically, starting from an assumption on the initial values of $l^0_{b,t}$ and $l^0_{g,t}$ we solve for $s_t$ using equation (54). Then we use equation (57) to solve for $\delta_t$ given that $LFP^0_t$ and $u_{0,t}$, and $u_{0,t+1}$ are observed. Next, we use (55) and (56) to compute $l^0_{b,t+1}$ and $l^0_{g,t+1}$, and repeat until we exhaust the sample.

To compute the inflationary effects of the estimated shocks to the rate of on-the-job search — shown in the right panel of Figure 10 — we need to solve the model nonlinearly. This is because inflation cannot be estimated recursively, as shown above for the labor market variables. To get around this problem, we combine the estimated shocks to the rate of on-the-jobs search with the impulse response function of inflation for the linearized model, which are shown in Figure 3. This approach gives us the red bar in the right plot of Figure 10.

# N A Model with Endogenous Search

One advantage of the model presented in Section 3 is that it allows us to derive a closed-form expression for an indicator of wage competition, and neatly show the role of its three driving forces: unemployment, labor misallocation, and the on-the-job search rate. An additional advantage is that this indicator can be derived from the observation of UE and EE transition, without having to solve the model. A disadvantage of that model, is that it relies on the assumption of an exogenous rate of on-the-job search, where all employed worker are assumed to look for jobs with the same probability, independently of their position on the ladder and independently of the state of the business cycle. Arguably, workers want to search more actively when employed in a bad match and during expansions, when the probability of getting an offer is relatively high and its expected surplus is larger. In this section, we show that endogenizing the rate of on-the-job search does not materially affect the estimated on-the-job search rate and the distribution of match quality, leaving our conclusions on the dynamics of labor market slack and missing inflation virtually unchanged.

We construct a model that is identical to the one studied in the previous sections, except that agents now optimally choose whether to search on the job or not. Specifically, each period, the employed worker $j$ draws an idiosyncratic fixed cost of search, $\varsigma_{j,t}$, from a uniform distribution

$$g(\varsigma_{j,t}) \sim \mathcal{U}\left[\xi_t\varsigma, \xi_t\varsigma + \varsigma\right], \tag{58}$$

where $\varsigma > 0$ is a parameter determining the support of the distribution and $\xi_t$ is an aggregate shock shifting the support of this distribution. The aggregate process affecting the cost of searching on the job is assumed to follow the AR process:

$$\xi_t = \left(1 - \rho_\xi\right) \bar{\xi} + \rho_\xi \xi_{t-1} + \varepsilon_{\xi,t}, \quad \varepsilon_{\xi,t} \backsim \boldsymbol{N}\left(0, \sigma_\xi\right), \tag{59}$$

where $\bar{\xi}$ is parameter capturing the unconditional mean of the process $\xi_t$. A negative realization of this shock shifts the support of the distribution downward, raising the probability of drawing a fixed cost of search that is lower than the expected return. Consequently, more agents would search on the job. While the shock, $\varepsilon_\xi$, affects the distribution of the fixed cost $\varsigma_{j,t}$ identically across workers, the individual response of a worker's propensity to search on the job depends on their position on the ladder and on the share of surplus they are able to extract from their current match.

The worker-specific costs $\varsigma_{j,t}$ are purely psychological and do not absorb households' resources. It should be noted that the lower bound of the support of the uniform distribution in equation (58) can take negative values, implying that employed workers may search even if they do not expect an increase in surplus. A negative shock should be interpreted as a psychological reward to changing jobs that is unrelated to its compensation and is unobservable by the firm. The aggregate shock $\xi_t$ is necessary to redo the empirical analysis of Section 4.5, which requires the model to have at least two shocks to jointly explain the two labor-market variables (the UE and EE rates). The other shock is the shock to households' preferences.

In this economy, every worker is willing to search on the job, provided that the stochastic cost of search, $\varsigma_{j,t}$, is below the expected return, i.e. the expected gain in surplus. The return to search will depend on $(i)$ the position of workers on the ladder, $(ii)$ the amount of match surplus that they are able to extract, and $(iii)$ the state of the business cycle, which affects both the probability of meeting a poaching firm and the expected surplus from being poached. The bargaining protocol—based on Bertrand wage competition among firms—is the same as in the benchmark model. In this setup, we need to track the decision to search for five types of workers: $(i)$ workers who are employed in a bad match and receive no surplus; $(ii)$ workers who are in a good match and receive no surplus; $(iii)$ workers who are in a bad match under full surplus extraction; $(iv)$ workers who are in a good match with partial extraction of surplus (i.e., they get the surplus of a bad match); $(v)$ workers who are employed with full extraction of a good-match surplus.[38] For each of the five categories of workers listed above, the fraction of workers who search on the job is determined by the uniform cumulative distribution; in symbols, $0 <\text{Prob}\{\varsigma_{j,t} < E_t \Delta S_t\left(k\right)\} = \frac{E_t \Delta S_t(k)}{\varsigma} - \xi_t < 1$, where $E_t \Delta S_t\left(k\right)$ denotes the expected surplus gain from on-the-job search for each of the five types of workers $(k = \{i, ii, iii, iv, v\})$. More details about the model are provided in Appendix O.

In this extended model, a positive preference shock, by increasing the probability of meeting

[38] Assuming that workers who are extracting the full surplus of their good match do not search would not materially affect the results of this robustness exercise.

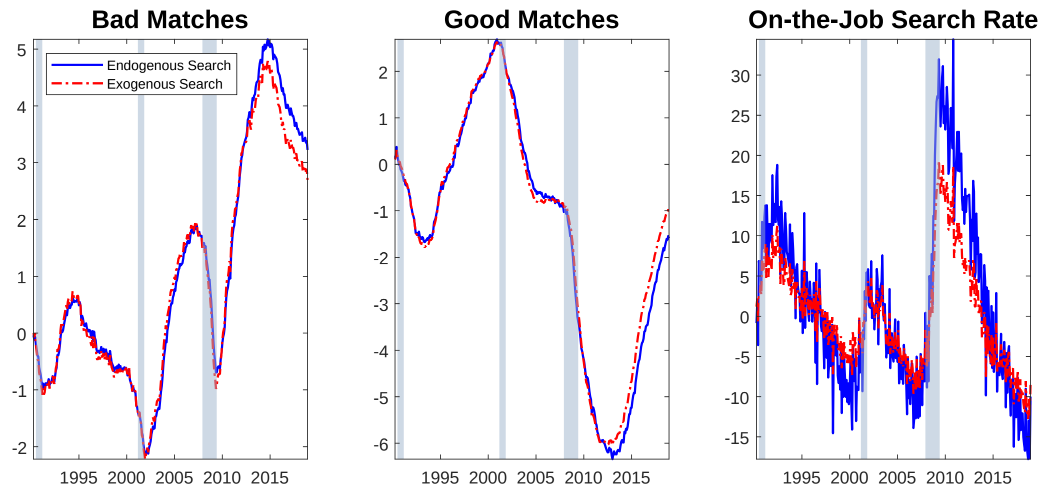


Figure 19: The in-sample dynamics of bad matches (left plot), good matches (central plot), and the on-the-job search rate (right plot) according to the benchmark model with exogenous on-the-job search (red dashed-dotted line) and the model with endogenous on-the-job search (blue solid line). Units: percentage points in deviation from their steady-state value.

a poaching firm and hence the expected return from on-the-job-search, increases the fraction of employed workers who look for jobs. This is a key departure from the benchmark model, in which the on-the-job search rate is exogenous and orthogonal to preference shocks.

To calibrate the model, we follow the same approach discussed in Section 4.1. This implies that all parameters that are common to the model with exogenous search will take the values reported in Table 1. There are only four parameters in the model with endogenous search that do not feature in the baseline model, and therefore need to be discussed here; they pertain to equations (58) and (59) above. The persistence and the standard deviation of the shocks to the distribution of the cost of on-the-job search, $\xi_t$, is estimated using a least-square approach.[39] The estimation leads to $\rho_\xi = 0.7524$ and $100\sigma_\xi = 5.08$. The parameters $\bar{\xi}$ and $\varsigma$ are set to $-0.1452$ and $0.5583$, respectively, so as to ensure that the steady-state rate of $EE$ transitions and the fraction of employed job seekers are consistent with the corresponding values in the model with exogenous on-the-job search.

Figure 19 compares the dynamics of bad matches, good matches, and the rate of on-the-job search (in percentage deviations from their steady-state value) obtained in the model with exogenous and endogenous on-the-job search (red dashed-dotted and blue solid lines, respectively). The two models deliver very similar results, with the endogenous-search model suggesting a somewhat more volatile on-the-job search rate and a somewhat larger slack at the end of the sample, as reflected in a larger mass of bad matches and a lower rate of on-the-job search rate.

[39] We make an initial guess on the parameters $\rho_\xi$ and $\sigma_\xi$ and then use the Kalman filter to obtain the estimated series for the stochastic process of the shocks $\tilde{\xi}_t \equiv (\xi_t - \bar{\xi})$ implied by the calibrated model and the data (the unemployment rate and the EE flow rate). We then regress the estimated series of $\tilde{\xi}_t$ on $\tilde{\xi}_{t-1}$ and obtain the OLS estimate of the parameters $\rho_\xi$ and $\sigma_\xi$, which are then used to check our guess. If these values diverge, we update our guess using the last OLS estimate of these parameters and redo the procedure until convergence.

Similarly to the benchmark model, the model with endogenous search accounts for the missing inflation of the last decade.

The reason why the two models provide similar predictions lies in the relative stability of the EE rate with respect to the UE rate. Because the distribution of workers along the ladder only changes slowly over time, the average rate of on-the-job search must account for the bulk of the changes in the EE to UE ratio, independently of whether the decision to search on the job is exogenous or endogenous. We relegate further details on the breakdown of search decisions across matches and surplus extraction to Appendix P.

# O The Equations of the Model with Endogenous Search

Assume that each period, every employed worker draws a fixed cost of search from a uniform distribution

$$g(\varsigma) \sim \mathcal{U}\left[\xi_t \varsigma, \xi_t \varsigma + \varsigma\right], \tag{60}$$

where $\varsigma > 0$ is a parameter. We assume that the aggregate shock to the cost of searching on the job behaves as follows

$$\xi_t = \left(1 - \rho_\xi\right)\bar{\xi} + \rho_\xi \xi_{t-1} + \varepsilon_{\xi,t}, \quad \varepsilon_{\xi,t} \backsim \boldsymbol{N}\left(0, \sigma_\xi\right). \tag{61}$$

Let $l_t^{i,u}$ denote the number of workers employed in matches of type $i \in (b, g)$ under zero surplus, and $\bar{\varsigma}_t^{i,u}$ the threshold value that makes them indifferent to search on the job or not. Similarly, let $l_t^{i,j}$ and $\bar{\varsigma}_t^{i,j}$ denote the measure of workers and threshold search costs that refer to workers employed in matches of type $i \in (b, g)$ under extraction of surplus of a job of type $j \in (b, g)$. The threshold value of the search cost that makes a worker employed in a bad match under zero surplus indifferent between searching on the job and not searching is $\bar{\varsigma}_t^{b,u} = \phi_t S_{b,t}$. For a worker employed in a good match under zero surplus, the threshold is $\bar{\varsigma}_t^{g,u} = \phi_t\left(\xi_b S_{b,t} + \xi_g S_{g,t}\right)$. For a worker employed in jobs of type $i$ under full extraction of a bad job surplus we get the following threshold: $\bar{\varsigma}_t^{b,b} = 0$. For worker employed in a good match under partial extraction, we obtain $\bar{\varsigma}_t^{g,b} = \phi_t \xi_g\left(S_{g,t} - S_{b,t}\right)$. Finally, for workers employed in good jobs under full extraction of the surplus, the threshold for searching is $\bar{\varsigma}_t^{g;g} = 0$.

The assumption that $g(\varsigma)$ is uniformly distributed implies that

$$Prob\left\{\varsigma_{j,t} < \bar{\varsigma}_t^{i,j}\right\} \equiv G\left(\varsigma < \bar{\varsigma}^{i,j}\right) = \frac{\bar{\varsigma}_t^{i,j}}{\varsigma} - \xi_t, \tag{62}$$

if $\xi_t \varsigma < \bar{\varsigma}_{j,t} < \xi_t \varsigma + \varsigma$. Note that $G\left(\varsigma < \bar{\varsigma}^{i,j}\right)$ for each type $(i,j)$ represents the fraction of workers whose period cost of search is lower than their threshold. By the law of the large numbers, this value equals the measure of workers of type $(i,j)$ who search on the job. These measures are necessary to characterize the workers' laws of motion across the rungs of the ladder.

The laws of motion for the workers employed under zero surplus is

$$l_t^{i,u} = l_{0,t}^{i,u} + \phi_t \xi_i u_{0,t} - l_{0,t}^{i,u} G\left(\bar{\varsigma}_t^{i,u}\right) \phi_t \quad \text{for } i = \{b, g\}, \tag{63}$$

$$l_{0,t+1}^{i,u} = (1-\delta)\, l_t^{i,u}. \tag{64}$$

The law of motion for the workers employed in bad matches under full extraction of bad match surplus is

$$l_t^{b,b} = l_{0,t}^{b,b} + l_{0,t}^{b,u} G\left(\bar{\varsigma}_t^{b,u}\right) \xi_b \phi_t - l_{0,t}^{b,b} G\left(\bar{\varsigma}_t^{b,b}\right) \phi_t \xi_g \tag{65}$$

$$l_{0,t+1}^{b,b} = (1-\delta)\, l_t^{b,b}.$$

The law of motion for the workers employed in good matches under full extraction of bad match surplus is

$$l_t^{g,b} = l_{0,t}^{g,b} + l_{0,t}^{b,u} G\left(\bar{\varsigma}_t^{b,u}\right) \xi_g \phi_t + l_{0,t}^{b,b} G\left(\bar{\varsigma}_t^{b,b}\right) \phi_t \xi_g + l_{0,t}^{g,u} G\left(\bar{\varsigma}_t^{g,u}\right) \xi_b \phi_t - l_{0,t}^{g,b} G\left(\bar{\varsigma}_t^{g,b}\right) \phi_t \xi_g \tag{66}$$

$$l_{0,t+1}^{g,b} = (1-\delta)\, l_t^{g,b}. \tag{67}$$

Equations (63), (65) and (66) above solve for $l_t^{i,u}$ and $l_t^{i,b}$, for $i = \{b, g\}$.

The total measure of workers searching at every point in time is:

$$l_{0,t}^{s} = l_{0,t}^{b,u} G\left(\bar{\varsigma}_t^{b,u}\right) + l_{0,t}^{g,u} G\left(\bar{\varsigma}_t^{g,u}\right) + l_{0,t}^{b,b} G\left(\bar{\varsigma}_t^{b,b}\right) + l_{0,t}^{g,b} G\left(\bar{\varsigma}_t^{g,b}\right). \tag{68}$$

The law of motion for the workers employed in good matches under full extraction of the good job surplus is:

$$l_t^{g,g} = l_{0,t}^{g,g} + \left[l_{0,t}^{g,u} G\left(\bar{\varsigma}_t^{g,u}\right) + l_{0,t}^{g,b} G\left(\bar{\varsigma}_t^{g,b}\right)\right] \xi_g \phi_t, \tag{69}$$

$$l_{0,t+1}^{g,g} = (1-\delta)\, l_t^{g,g}. \tag{70}$$

Total employment can be computed as

$$l_{0,t} = l_{0,t}^{b,u} + l_{0,t}^{b,b} + l_{0,t}^{g,u} + l_{0,t}^{g,b} + l_{0,t}^{g,g}. \tag{71}$$

# P The Evolution of Match-Quality in the Model with Endogenous Search

The shock $\xi_t$ is pinned down by the joint dynamics of the EE rate and the UE rate implied by the observed unemployment rate. The lower degree of procyclicality of the EE rate relative to the job finding rate leads the shock $\xi_t$ to lower the support of the distribution of search costs

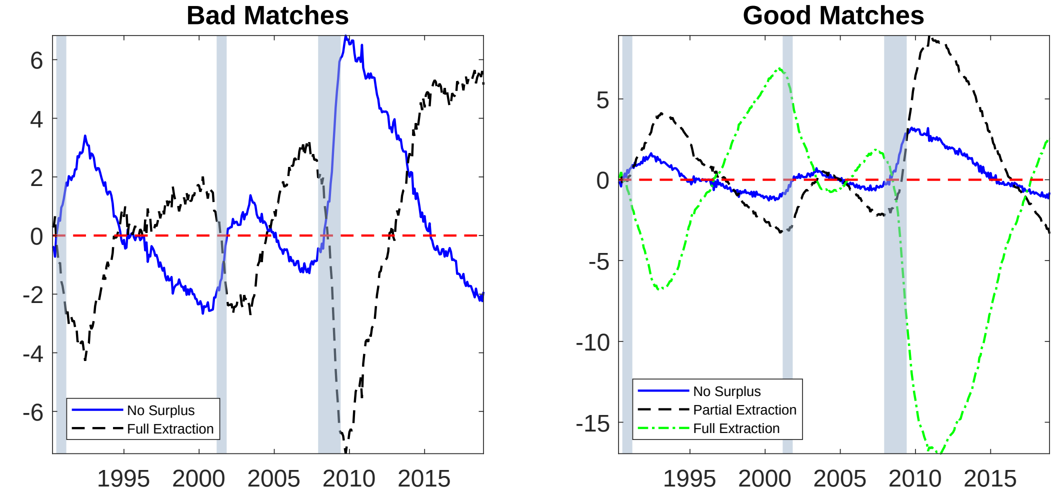


Figure 20: Breakdown of Bad Matches and Good Matches. Left plot: bad matches. Right plot: good matches. Unit: deviations from steady state in percentage points.

in recession and to raise it in expansion. A lower support implies that, everything else equal, more employed workers will draw search costs that lie below the threshold that makes them indifferent between searching or not. Hence, more of them will search. How the probability to search on the job responds to the shock $\xi_t$ depends on the position of workers on the ladder and the surplus they are able to extract. Specifically, the fraction workers employed in a bad match is very countercyclical and attains its historical low at the end of the sample.[40] This implies that the bad matches remain persistently above their long-run level throughout the last recovery. Remarkably, even though the job finding rate attains very high levels at the end of the sample, our model with endogenous on-the-job search predicts that the mass of bad matches decreases only very slowly. This sluggish adjustment in the measure of bad matches is mainly accounted for by the workers who are stuck in a bad match under full surplus extraction. A graph showing the breakdown in the behavior of bad matches by surplus extraction is provided in Figure 20.

The left plot of Figure 20 shows the decomposition of bad matches into workers in bad matches with no surplus (the blue solid line) and those with full surplus extraction of surplus (the black dashed line). The latter are workers in a bad match that were able to secure an outside offer from another firm that is a bad match for the worker. As one can see, the share of these workers raises at the beginning of the expansion, which is typical of the job ladder, but, instead of converging back to steady state, it levels out until the end of the expansion.

[40]Workers who have not received any outside offers and thereby receive no surplus are the ones who search more intensively. However, their propensity to search is highly procyclical being primarily affected by the dynamics of the job finding rate. At the end of the sample, these workers search a lot A similarly procyclical pattern is followed by the search rate of those workers employed in good matches under partial extraction of surplus (i.e., they are in a good match but they have so far failed to secure a good outside offer). Nevertheless, their search rate is less volatile than those of the workers who are extracting zero surplus.

This happens because the propensity of these workers to search on the job declines during the recovery.

The presence of workers who are stuck in a bad match, having failed to attract a good offer, contributes to mitigate the intensity of interfirm wage competition. As explained in Section 3.6 in the context of the baseline model, these workers represent a relatively cheap source of labor from the perspective of an entrant firm, given their expected productivity. As a result, this large share of bad matches brings about slack, thereby exerting downward pressure on inflation. Indeed, the model with endogenous on-the-job search predicts inflation to remain below its longer-run level (2%) throughout the long post-Great Recession recovery, echoing the results obtained in the baseline model in Section 4.

# Q Macroeconomic Identification of on-the-Job Search Rates: a Discussion

In our model, EE flows in equation (20) can be rewritten as:

$$EE_t = s_t \phi_t^U \times P(M_t), \tag{72}$$

where $s_t$ is the probability of searching on the job, $\phi_t^U$ is the job finding rate from unemployment, and $P(M_t)$ stands for the probability that employed workers switch their jobs conditional on receiving an offer. This probability depends on the distribution of jobs by match quality $(M_t)$ - i.e., the share of good and bad matches $l_{g,t}^0$ and $l_{b,t}^0$- at the beginning of time $t$. This equation stems from the assumption that conditional on searching with probability $s_t$, a worker who is employed finds a job with the same probability of a worker who is unemployed, $\phi_t^U$. The distribution of workers across the ladder, encoded in the function $P(M_t)$, will also contribute to determine the rate at which job offers to the employed are actually converted into jobs. The specification in equation (72) is admittedly an overly simplified representation of the many factors that actually drive EE transitions. A more accurate description of the process would be:

$$EE_t = s_t^{ext} s_t^{int} w_t \phi_t^U \times P(M_t),$$

where EE transitions are now assumed to depend on whether an employed worker looks for jobs—search on the extensive margin takes place with probability $s_t^{ext}$—, but also how much time the worker spends on searching—$s_t^{int}$ would be a measure of the intensive margin of search. One could also define $w_t \phi_t^U \equiv \phi_t^E$ and interpret $w_t$ as a time-varying wedge between the job finding rates of the employed, $\phi_t^E$, and that of the unemployed, $\phi_t^U$. Changes in this wedge would pick up time-variation in the composition of the pool of both the employed and the unemployed job seekers as well as changes in firms' preferences for hiring employed vs. unemployed workers.

Our empirical approach allows us to measure $s_t$, but what the estimated $s_t$ is really picking up is $s_t^{ext} s_t^{int} w_t$. How does our estimated series for $s_t$ based on equation (72) compare to estimates of $s_t^{ext}$ based on the survey data? To address this question, we explore a new survey, which has been administered by the Federal Reserve Bank of New York as a supplement to the *Survey of Consumer Expectations* (SCE). The SCE is a monthly and nationally representative survey of about 1,300 individuals, which directly asks employed workers whether they have been actively searching for work in the previous seven days.[41] In this paper, we use SCE data available from 2014Q4 trough 2019Q4. Even if this is admittedly a short period of time, it still covers the years in which the fall in the rate of on-the-job search predicted by our model is critical to account for the missing inflation. Figure 7 plots the on-the-job search rate implied by the model, $s_t$, and the corresponding measure in the survey data (the blue solid line and the black line with stars, respectively). The figure shows that the fall in the on-the-job search rate predicted by our model using aggregate labor market flows is strikingly close to the one measured in the survey data.

Using information on the hours of search for the employed workers in the SCE, we find that the fall in the aggregate amount of time spent searching is entirely explained by the extensive margin; that is, the effect is due to a fall in the incidence of job search among the employed—and not to a decrease in the average number of hours dedicated to search. We also looked at how the arrival rate of job offers for the employed workers varied over our sample period, relative to the arrival rate of offers for the unemployed. The evidence does not indicate a divergence in the arrival rate of offers for the employed and the unemployed.[42]

While it is interesting to notice that from 2014 and onwards the estimated pattern of $s_t$ nearly coincides with the share of employed workers searching on the job, as measured in the survey data, we cannot rule out that other factors may also have been at play, particularly around the times of the Great Recession. What is important to notice though, is that for the question we are after, which is how a job ladder model can explain the missing inflation at the end of the previous decade, it does not really matter whether the variation that we extract in $s_t$ accurately reflects changes in the extensive margin of search, or rather picks up other components that could be assigned to $s_t^{int}$ or $w_t$. What matters for inflation is that the employed workers become less effective in obtaining outside offers, thereby decreasing interfirm wage competition. Whether this is due to a fall in the extensive margin of on-the-job search, a fall in the intensive margin, a change in the composition of the employment pool that makes

[41] Question JS9 of the survey asks the following: "And within the LAST 7 DAYS, about how many TOTAL hours did you spend on job search activities? Please round up to the nearest total number of hours." We drop self-employed workers when computing the on-the-job search rate from the SCE.

[42] We computed, both for the employed and the unemployed, the ratio between the total number of offers received—and not necessarily accepted—and the aggregate total number of hours spent searching.

them less inclined to search or receive offers, or a change in hiring practices that makes firms less inclined to propose job offers to the employed, it does not really matter for the object of our investigation.