Federal Reserve Bank of Chicago

# Bad Jobs and Low Inflation

*Renato Faccini and Leonardo Melosi*

March 2, 2020

WP 2020-09

https://doi.org/10.21033/wp-2020-09

# Bad Jobs and Low Inflation*


Renato Faccini
Danmarks Nationalbank
Queen Mary, University of London
Centre for Macroeconomics (LSE)

Leonardo Melosi
FRB Chicago
European University Institute
CEPR


March 4, 2020


**Abstract**

We study a model in which firms compete to retain and attract workers searching on the job. A drop in the rate of on-the-job search makes such wage competition less likely, reducing expected labor costs and lowering inflation. This model explains why inflation has remained subdued over the last decade, which is a conundrum for general equilibrium models and Phillips curves. Key to this success is the observed slowdown in the recovery of the employment-to-employment transition rate in the last five years, which is interpreted by the model as a decline in the share of employed workers searching for a job. This fall in the on-the-job search rate is corroborated by the micro data.



*Keywords*: Missing inflation, job ladder, cyclical misallocation, labor market slack, Phillips curve.

*JEL codes*: E31, E24, C78



*Correspondence to: r.faccini@qmul.ac.uk and lmelosi@frbchi.org.


We thank Gadi Barlevi, Robert Barsky, Marco Bassetto, Lawrence Christiano, Martin Eichenbaum, Jason Faberman, Michael Krause, Giuseppe Moscarini, Fabien Postel-Vinay, and seminar participants at the Society for Economic Dynamics, Chicago Fed, European University Institute, LSE, Northwestern University, Bank of England, 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. May Tysinger provides 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 other person associated with the Federal Reserve System or the European System of Central Banks.

# 1 Introduction

Workhorse models 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 recent U.S. macroeconomic developments if the rate of unemployment is taken as a proxy for labor market slack, following a conventional approach dating back to Phillips (1957). As shown in Figure 1, since March 2017 the unemployment rate has reached its 50-year low at 3.5%, hovering consistently below its average level measured over the last twelve months of the previous expansion. At the same time, PCE core inflation has remained persistently below its long-term expectations. We first show that standard measures of labor market slack fail to explain the missing inflation. We then show that a general equilibrium model, whose relevant measure of slack is influenced by the fraction of workers searching on the job, can explain the lack of inflationary pressures 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 decline in the 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 rise 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 rate and the employment-to-employment (EE) flow rate observed in the data. The EE rate has slowly recovered during the expansion and has leveled out since 2015. A low and stagnant fraction of workers who switch jobs, combined with a tight labor market in which finding a job becomes easier over time, are interpreted by our model as a fall in the on-the-job search rate. To validate this prediction, we estimate the on-the-job search rate at the micro level using the *Survey of Consumer Expectations* (SCE) administered by the Federal Reserve Bank of New York. In the *Survey*, the on-the-job search rate has fallen from 2014 through 2017 in a way that is remarkably similar to our measure based on the aggregate labor market flows.

We derive a model-consistent concept of labor market slack, which can be measured using the observed series of the unemployment rate and the EE rate. Labor market slack hinges on 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. A fall in the fraction of workers who are searching on the job lowers labor market slack in the model because it increases the firms' chances to fill their

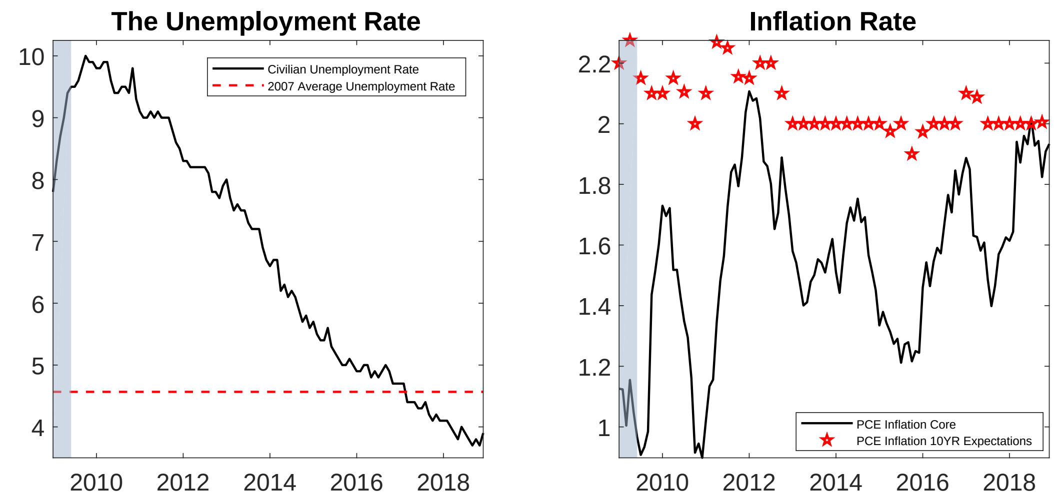


Figure 1: Labor market and inflation dynamics during the post-Great Recession recovery. The left graph: the civilian unemployment rate (solid line) and its average computed over the twelve months that preceded the Great Recession (red dashed), computed over the period December 2006 through November 2007. Frequency: Monthly. Source: Bureau of Labor Statistics (BLS). The right graph: core PCE inflation (annualized in percentage, black line). Frequency: Monthly. Source: BEA. The red stars denote the 10-year ahead expectations about the PCE inflation rate. Frequency: Quarterly. Source: The Federal Reserve Bank of Philadelphia - the Survey of Professional Forecasters. The shaded areas denote NBER recessions.

vacancies with unemployed workers, who are cheaper to hire as the unemployed workers are unable to prompt wage competition between employers. The more inefficient is the allocation of labor, the more likely it is for firms to meet workers employed in low-productivity (bad) matches. Since 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 raises labor market slack.

We then take the model to the unemployment rate and the flow EE rate 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 sole purpose of generating the fluctuations in the unemployment rate observed in the data. Given the unemployment rate, the EE rate allows us to pin down the shocks to the on-the-job search rate as described earlier. Equipped with the time series for the two shocks, we simulate the path of inflation and labor costs implied by the model. We find that the model does not see inflationary pressures during the current expansion and this result is driven by the decline in the on-the-job search rate which has kept wage competition at low levels. In addition, labor-cost dynamics in the model closely correlate with the growth rate of the average hourly earnings and the employment cost index.

When we analyze the contribution of each of the three components of labor market slack to inflation during the post-Great Recession period, we find that the drop in the on-the-job search rate emerges as the key explanation for why inflation is still so low in the U.S. after nine years of economic recovery. Labor misallocation also contributes significantly to keeping inflation persistently subdued following the Great Recession, offsetting the effects of the low unemployment rate.

The surge in labor misallocation right after the recession is due to the exceptionally high stock of unemployed workers who took a first step back onto the job ladder. As a result of the persistent decline in the on-the-job search rate throughout the recovery, the speed at which workers moved to better jobs fell, exacerbating labor misallocation and exerting persistent downward pressures on wage dynamics and inflation. Indeed, our model predicts that after nine years of expansion, a significant fraction of the employed workers is still stuck in suboptimal jobs. This prediction is consistent with the microevidence from the SCE, which shows that about 30 percent of workers are not fully satisfied with their current occupation in 2017. This persistent rise in bad jobs also accords well with evidence in Jaimovich et al. (2020), who show that a third of the workers that were employed in routinary occupations before the Great Recession could not find similar jobs and are now stuck in nonroutinary manual occupations.

In the post-war period, the U.S. economy has experienced low rates of unemployment and inflation in other circumstances, like in the 1960s and in the 1990s. However, these episodes occurred in connection with high labor productivity growth, which in New Keynesian models lowers real marginal costs and hence dampens inflationary pressures. What makes the current expansion so puzzling is that inflation has remained low while labor productivity growth has also slowed down (Fernald 2016). By predicting a persistent surge in the incidence of low-productivity jobs in the recent recovery, our model reconciles the absence of inflationary pressures with a dismal labor productivity growth.

How does the model fare in the earlier period? To address this question, we compare the performance of our model-consistent measure of labor market slack to that of other popular measures in the literature, such as the labor share of income (as in Galí and Gertler 2000), the unemployment gap based on the NAIRU, and the hours worked, which features prominently in structural estimation of dynamic general equilibrium models as the key observable variable informing the output gap (e.g., Christiano, Eichenbaum, and Evans 2005). We find that in this earlier sample period (1990 through 2012), our measure of slack performs comparably to these other popular measures while it does significantly better at accounting for the missing inflation in the last decade.

The result that inflation has been moderate in the current recovery relies on the countercyclicality of the search rate. When we extend the analysis back to the early 1990s, we find that the search rate has been countercylical in this earlier period too. We show that such countercyclicality stems from the fact that the volatility of the unemployment rate, which in the model reflects the probability of finding a job, is higher than the volatility of the EE flow rate in the data. We elaborate on the reasons why the on-the-job search rate is countercyclical by connecting to a number of findings in the empirical micro-labor literature.

Assuming that the on-the-job search rate varies stochastically over time is meant to capture all those cyclical factors that drive the decision to search on the job, as well as compositional changes in the propensity to search within the pool of employed workers.[1] We do not explicitly model these compositional changes in our macro model and assume the on-the job search is exogenous. We believe that this is the right approach at this stage as no consensus about how this rate varies over the business cycle has been reached yet in the micro-labor literature. In addition, since the time series of the on-the-job search rate is uniquely pinned down by observing the unemployment rate and the EE flows, endogenizing this rate could change our results only by affecting agents' expectations about the future evolution of the rate. We show that this expectation channel is not strong enough to alter our main conclusions about the model's ability to explain the missing inflation.

Our model features an occasionally binding zero lower bound (ZLB) constraint on the nominal interest rates. Introducing this constraint is important since the severity of the Great Recession, which in our analysis is captured by the sharp increase of the unemployment rate in 2008 and 2009, drives the current and expected nominal interest rates to the ZLB for several months in our model. We develop an innovative method to solve and simulate models when the ZLB constraint is binding. Our method does not rely on assuming perfect foresight.

Moscarini and Postel-Vinay (2019) pioneer a New Keynesian model in which cyclical labor market misallocation brings about deflationary pressures. In building our model, we draw from their path-breaking theoretical contribution. These scholars use the model to argue 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 ways. First, while their empirical analysis is reduced form and external to their structural model, we take our structural model to the data using state-of-the-art time series methods. Second, we allow the on-the-job search rate to vary over time and this is key to explain the missing inflation of the last decade. In Moscarini and Postel-Vinay's model, the search rate is constant implying that the acceptance ratio, which is the ratio of EE to UE flow rates, is a leading indicator for inflationary pressures. This ratio is a proxy for the degree of cyclical labor misallocation and in their model a low value of this ratio predicts high inflation.[2] Currently the acceptance ratio is lower than its pre-Great Recession average in the data, as shown in Appendix A. Our model jointly explains this low acceptance ratio, 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. To put it bluntly, according to our model the acceptance ratio is currently low in the data, not because employment is efficiently allocated but because less workers are searching on the job.

[1] For instance, workers who are hired at the beginning of an expansion may be more dynamic than those who generally find jobs when the labor market is already very tight (Cahuc, Postel-Vinay, and Robin 2006).

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

Understanding the search behavior of the employed using disaggregated labor data is an active area of ongoing research. In this paper we show that this line of research is also important to better assess the degree of labor market slack and inflationary pressures in the economy. Abraham and Haltiwanger (2019) survey this literature and analyze the properties of a novel measure of labor market tightness relative to those of the standard series. Our measure of slack differs from theirs insofar as we propose a theory-based measure that relies on EE flow data for the identification of the rate of the on-the-job search. Moreover, these scholars do not attempt to econometrically evaluate how well their measure of slack explains inflation.

The paper is related to the empirical literature that studies the role of search and matching frictions in New Keynesian models. Key empirical studies include Gertler, Sala, and Trigari (2008), Krause, Lopez-Salido, and Lubik (2008), 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.

The paper is organized as follows. In Section 2 we motivate the paper by showing the missing inflation puzzle. The general equilibrium model is introduced in Section 3, while we deal with the empirical analysis in Section 4. Section 5 presents the conclusions.

# 2 The Existing Theories of Inflation

The New Keynesian model is the most popular framework to study inflation. A key building block of the New Keynesian framework is the New Keynesian Phillips curve, which posits that inflation $\pi_t$ hinges on the expected dynamics of 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), which link inflation to the current and expected unemployment gap; $(ii)$ the standard New Keynesian theory, derived from models with no labor frictions, suggesting 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

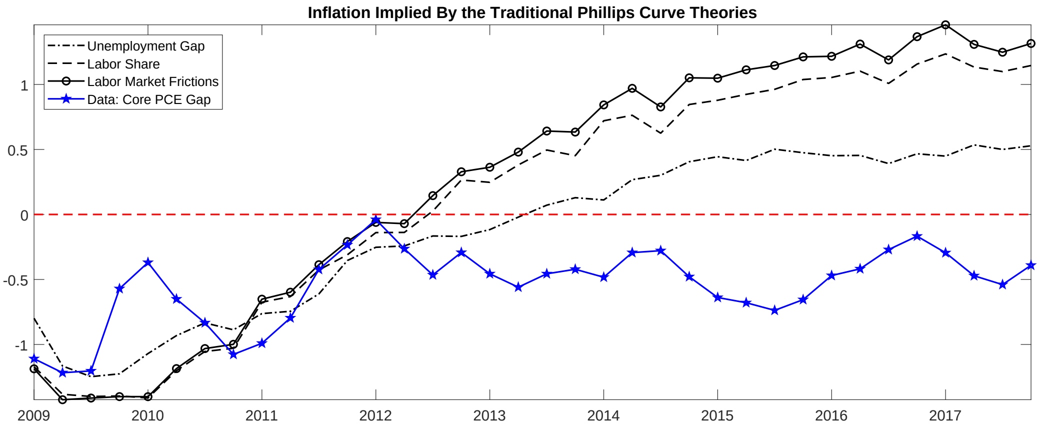


Figure 2: Core PCE inflation gap from 2009Q1 through 2017Q3 and inflation dynamics predicted using three traditional theories of inflation.

measures of the labor share as well as UE flow rates (Krause, Lopez-Salido, and Lubik, 2008).[3] 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 the main conclusions of this exercise.

Solving equation (1) forward, expected inflation can be expressed 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 is 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. While the slope of the Phillips curve affects the magnitude of inflation predicted by the three theories, it does not affect significantly the point in time when inflation rises above target, which is what we are interested in.

To estimate the VAR model, we construct a gap measure for the quarterly macro observables 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.[4] We rely on the NAIRU estimates to construct the unemployment gap as this practice is very popular in those studies

[3] 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.

[4] Using the long-term NAIRU estimates to construct the unemployment gap does not change the results significantly.

whose object is to estimate the Phillips curve. The observables are: the labor share, the job finding rate, real wages, the civilian unemployment rate, real GDP, real consumption, real investment, CPI inflation, and the federal funds rate (FFR).[5] The data sample covers the period 1958q4 through 2017Q4. The VAR model is estimated using the gap of these nine observables, and the forecasts of the three measures of marginal costs are launched at every quarter starting from the first quarter of 2009 through the fourth quarter of 2017.

Figure 2 shows that all the three traditional theories of marginal costs suggest that inflation should have been above its long-run level (positive inflation gap) starting around 2013 or 2014. None of these theories is able to account for why inflation has been so low for so many years after the Great Recession because all the 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.

This VAR-based approach does not require us to take a stand on what model people use to form their expectations about future labor costs in the real world. Imposing such a model may lead to misspecification that could distort our results. Our approach mitigates this problem as VAR models are reduced-form, theory-free representations for the data that are less prone to misspecification than structural theory-based models. For example, if we parametrically restrict the VAR model so as to make inflation to behave consistently with the Phillips curve, we most likely worsen the quality of the forecasts of real marginal costs. Our approach is general and agnostic as we do not impose that people use the Phillips curve when forming expectations in the real world. Moreover, another related advantage of our approach is that large Bayesian VAR models generally provide reliable macroeconomic forecasts.[6]

# 3 A General Equilibrium Model with the Job Ladder

The limitation of the traditional theories of inflation discussed above motivates the need of an alternative theory, which will be introduced in this section. The mechanism that we propose builds on the traditional New Keynesian paradigm but it shifts the emphasis on the role of time-variation in the wage competition for employed workers induced by changes in the incidence of

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

[6] For the Bayesian VAR model to deliver reliable macro forecasts, the choise of the prior is key. We check that VAR forecasts are accurate in sample and follow the conventions established by 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 of the VAR model.

on-the-job search.

## 3.1 The Economy

The economy is populated by a representative, infinitely lived household, whose members' labor market status is either unemployed or employed. All members of the households are assumed to pool their income at the end of each period and thereby consume the same. The labor market is frictional and workers search for jobs both whether they are unemployed or employed. While all unemployed workers are also job seekers, it is assumed that any employed worker can search in a given period with a probability $s_t$, which is assumed to follow an exogenous 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 who match in a frictional labor market and produce a homogeneous non-storable 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 $\varphi_t$ (the price of the labor service relative to that of the numeraire), 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 non-negativity 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 which 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, $1 - u_{0,t}$. Denoting the aggregate mass of vacancies by $v_t$, we can define labor market tightness as:

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

We assume that the meeting function is homotetic, which implies that the rate at which searching workers locate a vacancy, $\phi(\theta) \in [0,1]$, and the rate at which vacancies locate 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 employer. 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

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

### 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. All workers share their consumption risk within the households, allowing us to solve the problems from the perspective of a representative household.

**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 let $U(C_t)$ denote the utility function and $\mu_t$ 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 of firms that produce differentiated goods, $D_t^P$, profits of 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 attention to equilibria where the value of being employed for any worker is no less than the value of being unemployed. In this set-up, 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 separation shock is realized, and zero otherwise. The intertemporal maximization problem reads:

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

subject to the budget constraint,

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

the stochastic process for the employment status,

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

and for equilibrium wages $w_t(j)$.[7]

[7] The evolution of individual wages must obey 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, as we will

Equation (6) implies that a worker 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 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

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

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

$$\int_0^1 G\left(q_t(i)/Q_t\right) di = 1, \tag{7}$$

which nests Dixit-Stiglitz as a special case. 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. Relative to Dixit-Stiglitz, the Kimball aggregator introduces more strategic complementarity in price setting, which causes firms to adjust prices by less to a given change in marginal costs. 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 $G\left(q_t(i)/Q_t\right)$:

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

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 a demand function for the differentiated good $(i)$:

$$\frac{q_t(i)}{Q_t} = \frac{1}{1+\varkappa}\left(\frac{P_t(i)}{P_t \Xi_t}\right)^{\iota} + \frac{\varkappa}{1+\varkappa}, \tag{8}$$

where $\varkappa \leq 0$ is a parameter, $\iota = \frac{\chi(1+\varkappa)}{1-\chi}$, and $\Xi_t$ is the Lagrange multiplier associated with

show later. However, there are many streams of wages that can deliver the promised present discounted value of utility, making the distribution of the individual wages indeterminate. It can be shown that this indeterminacy is inconsequential for aggregate equilibrium outcomes. Nevertheless, 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 if the underlying wage distribution is not.

the constraint (7) and the aggregate price index (i.e., the price of the numeraire) satisfies $1 = \int_0^1 \left(\frac{p_{t,i}}{P_t \Xi_t}\right)^{\frac{\iota}{\omega^k}} di$.

## 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 don't get a chance to reoptimize, they adjust their price at the steady state inflation rate $\Pi$. Therefore, the problem of the price setting firm is:

$$\max_{P_{t+s}(i)} E_t \sum_{s=0}^{\infty} \beta^{t+s} \zeta^s \frac{\lambda_{t+s}}{\lambda_t} \left(P_t(i)\Pi^s - P_{t+s}\varphi_{t+s}\right) q_{t+s}(i) \tag{9}$$

subject to the demand function (8). Loglinearization 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; that is, price setters that are allowed to re-optimize their price at time $t$ find it optimal to forecasts 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 firm and discuss the pivotal role played by this condition in determining the dynamics of price setters' marginal costs and inflation in the model. 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 first discuss the expected costs incurred by entrant service sector firms, and 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 is 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 and 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 vacancy meets an employed worker instead, it engages in Bertrand competition with the incumbent firm in an attempt of poaching the worker. An important implication of these assumptions is that an increase in wages is not necessarily backed by a rise in workers' productivity. This can happen, for instance, when workers renegotiate upwards the value of their contract, as their employer agrees to match the offer of a poaching firm. This temporary decoupling between wages and workers' productivity is key for the job ladder to have meaningful implications for inflation. As we will show, these assumptions also imply that the worker's ability of extracting more and more surplus from a match depends on her position on the job ladder.

While the assumption that unemployed workers have no bargaining power is undoubtedly stark, 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 to provide intuition about the link between labor market variables and inflation in the model, which will be the focus of Section 3.6. This model-consistent measure of slack can be directly measured in the data.

Let $y$ and $y'$ denote 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 unemployment, she appropriates the surplus $S_t(y_b)$ since 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 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 job or staying. We assume that switching takes place with probability $\nu$ (a non-zero value for this parameter is required to match the high churning rate in the U.S. market when calibrating the model's steady-state parameters). In either case, the firm that retains 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 non-zero 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:

$$
\begin{aligned}
c^f + \frac{c}{\varpi_t} &= \frac{u_{0,t}}{u_{0,t} + s_t\left(1 - u_{0,t}\right)} \left\{\xi_b S_t\left(y_b\right) + \xi_g S_t\left(y_g\right)\right\} \\
&\quad + \frac{s_t\left(1 - u_{0,t}\right)}{u_{0,t} + s_t\left(1 - u_{0,t}\right)} \left\{\xi_g \frac{l^0_{b,t}}{1 - u_{0,t}} \left[S_t\left(y_g\right) - S_t\left(y_b\right)\right]\right\},
\end{aligned} \tag{10}
$$

where $l^0_{b,t}$ denotes the measure of workers who, at the beginning of period $t$, are employed in low quality matches ($l^0_{b,t} + l^0_{g,t} + u_{0,t} = 1$) and $s_t$ is the on-the-job search rate. It should be noted that $s_t\left(1 - u_{0,t}\right)$ denotes the measure of employed workers searching on the job at the beginning of period $t$ and $u_{0,t} + s_t\left(1 - u_{0,t}\right)$ 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 match with the incumbent. Three contingencies will give a nonzero surplus to the firm and will hence appear in the right-hand side of the free entry equation (10). The expected return on the right hand side is just an 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 $u_{0,t}/[u_{0,t} + s_t\left(1 - u_{0,t}\right)]$, 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 $S_t\left(y_b\right)$. 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 $S_t\left(y_g\right)$. These two expected returns appear in the first term on the right-hand side of the free-entry equation (10). The third contingency, i.e. the second term in the right-hand side of the free entry equation (10), occurs when the firm meets an employed worker, with probability $s_t\left(1 - u_{0,t}\right)/\left[u_{0,t} + s_t\left(1 - u_{0,t}\right)\right]$, and the following two conditions are met: (*i*) the worker is a good match for the entrant firm, which happens with probability $\xi_g$, and (*ii*) is currently in a bad match, which occurs with probability $\xi_g l^0_{b,t}/\left(1 - u_{0,t}\right)$.[8] As explained above, this is the only case in which an entrant firm can extract a nonzero surplus from meeting with an employed worker.

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

$$
S_t\left(y\right) = y\mathcal{W}_t - \frac{b\lambda_t^{-1}}{1 - \beta\left(1 - \delta\right)}, \tag{11}
$$

[8] 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 $\left(1 - u_{0,t}\right)$ so as to obtain the conditional probability of meeting a bad match.

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

$$\mathcal{W}_t = \varphi_t + \beta \left(1 - \delta\right) E_t \frac{\lambda_{t+1}}{\lambda_t} \mathcal{W}_{t+1}. \tag{12}$$

See Appendix F for details on the derivations in the context of our model. From the point of view of a labor service 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, who 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.

## 3.6 Wage Competition, Labor Market Slack, and Inflation

To provide intuition on how the labor market affects inflation dynamics in the model, we inspect the free-entry condition (10) and the Phillips curve (1) in isolation (i.e., in partial equilibrium). Let us assume that employed workers search less frequently ($s_t$ decreases). The direct effect of this change on the free-entry condition is to increase the likelihood for entrant firms to meet an unemployed job seeker. Since now it is more likely for firms to meet a worker they can extract a positive surplus from, the expected profits (i.e., the right hand side of the free-entry condition) will increase. As a result, more firms want to enter the labor service sector, enticed by the expected gains from posting a vacancy.

Two variables adjust to restore equilibrium in the free-entry condition. On the one hand, the increase in vacancies leads to a fall in the vacancy filling rate and to an increase in the expected cost of entry (i.e., the left hand side of equation(10)). On the other hand, the expected discounted stream of the relative price of labor service $\mathcal{W}_t$ falls, lowering surpluses $S_t\left(y\right)$ as implied by equations (11) and (12) and further contributing to dissuading firms from posting new vacancies until the free-entry condition is satisfied. This drop in the relative price of the labor service causes price setters' marginal costs to fall, lowering inflation.[9]

Inflation becomes lower because forward-looking price setters anticipate that over the current and future periods, wage competition to hire a worker will be less likely in the labor service sector. Indeed, a fall in the rate of on-the-job search increases the likelihood that a labor service firm will meet an unemployed worker. In such a case, the firm does not have to compete with another firm for the worker by bidding up its wage offer. Consequently, price setters will expect that the future prices of the labor service will fall for a while, prompting the price setters that can re-optimize to lower their price. As a result, the inflation rate falls.

[9] The present discounted stream of real marginal costs $\mathcal{W}_t$ falls only in the presence of nominal rigidities. 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.

A larger fraction of workers employed in bad matches also contributes to keeping wage competition low. When an entrant firm finds a good match in a worker who is already employed in a bad match, wage competition to attract this worker is not so intense to prevent the poaching firm from gaining a positive share of match surplus. A larger share of bad matches, everything else equal, decreases the expected intensity of interfirm wage competition, leading price setters to expect a shallower path of real marginal costs and hence inflation to fall.

This partial-equilibrium analysis suggests 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 is the key predictor of labor costs and inflationary pressures in the model. This probability is defined as follows:

$$\Sigma_t \equiv \frac{u_{0,t}}{u_{0,t} + s_t (1 - u_{0,t})} + \frac{s_t (1 - u_{0,t})}{u_{0,t} + s_t (1 - u_{0,t})} \xi_g \frac{l^0_{b,t}}{1 - u_{0,t}}, \tag{13}$$

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 who is employed in a bad match, is searching on the job, and is a good match for the poaching firm.

As we explained in this partial-equilibrium experiment, a high value of this probability leads price setters to expect a low intensity of wage competition and hence to anticipate falling marginal costs, which ultimately cause the rate of inflation to decline. Hence, this probability can be thought of as a measure of labor market slack. Indeed, as we will show, the link between the probability in equation (13) and inflation is very strong in our general equilibrium model.

The notion of labor market slack provided in equation (13) has two main advantages. First, as we will show in Section 4.4, it allows us to decompose inflation into its three drivers: the unemployment rate, $u_{0,t}$, the measure of bad jobs, $l^0_{b,t}$, and the on-the-job search rate, $s_t$. Such a decomposition will turn out to be useful to isolate the quantitative contribution of each of these three variables to the missing inflation of the last decade. Second, this measure of slack can be directly computed from the observed unemployment and EE flow rates with no need to solve the model, as it will be shown in Section 4.3.

## 3.7 The Dynamic Distribution of Match Types

The laws of motion for bad and good matches are

$$l_{b,t} = \left[1 - s_t \phi(\theta_t) \xi_g\right] l^0_{b,t} + \phi(\theta_t) \xi_b u_{0,t}, \tag{14}$$

$$l_{g,t} = l^0_{g,t} + s_t \phi(\theta_t) \xi_g l^0_{b,t} + \phi(\theta_t) \xi_g u_{0,t}. \tag{15}$$

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 $l_{b,t}^{0}$ and $l_{g,t}^{0}$ denote beginning-of-period values, instead. In turn, $l_{b,t}$ is equal to the sum of the bad matches at the beginning of a period who did not move up the ladder by finding a good quality match within the period, $\left[1 - s_t \phi\left(\theta_t\right) \xi_g\right] l_{b,t}^{0}$, 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 who search on the job with exogenous probability $s_t$, meet a vacancy with probability $\phi\left(\theta_t\right)$ and draw a good quality 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 $l_{g,t}^{0}$, plus the job-to-job inflows from low quality matches $s_t \phi\left(\theta_t\right) \xi_g l_{b,t}^{0}$, and the new hires from unemployment $\phi\left(\theta_t\right) \xi_g u_{0,t}$. Using the identity $l_{i,t+1}^{0}\left(y\right) = \left(1-\delta\right) l_{i,t}\left(y\right)$ for $i = \{b, g\}$ ,we can rewrite the dynamic equations (14) and (15) to express the laws of motion for bad and good jobs at their beginning-of-period values:

$$
\begin{aligned}
l_{b,t+1}^{0} &= (1-\delta)\left\{\left[1 - s_t \phi\left(\theta_t\right) \xi_g\right] l_{b,t}^{0} + \phi\left(\theta_t\right) \xi_b u_{0,t}\right\}, \qquad (16) \\
l_{g,t+1}^{0} &= (1-\delta)\left\{l_{g,t}^{0} + s_t \phi\left(\theta_t\right) \xi_g l_{b,t}^{0} + \phi\left(\theta_t\right) \xi_g u_{0,t}\right\}. \qquad (17)
\end{aligned}
$$

## 3.8 Policymakers and Market Clearing

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

$$
\frac{R_t}{R^*} = \max\left\{\frac{1}{R^*}, \left(\frac{R_{t-1}}{R^*}\right)^{\rho_r}\left[\left(\frac{\Pi_t}{\Pi^*}\right)^{\phi_\pi}\left(\frac{Q_t}{Q^*}\right)^{\phi_y}\right]^{1-\rho_r}\right\}, \qquad (18)
$$

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 (18) because these shocks cannot be separately identified by preference shocks in our empirical analysis. Indeed, the unemployment rate and the EE flow rate, which are the observables, respond very similarly to these two shocks.[10] To disentangle these two shocks, one has to add some other series; e.g., the nominal interest rate. However, adding nominal variables is undesirable as these variables could indirectly give our model information about the inflation rate. Instead, our empirical analysis about the ability of the model to explain inflation in the last decade is conditioned solely on *real* labor market variables. We consider this an important feature of our analysis.

[10] We note a fair amount of cannibalization between these two shocks when monetary shocks are added to the analysis. As a result, our main results would not change.

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 (8). 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 we examine the propagation of the shocks to preferences and search intensity. Section 4.3 explains how we implement our empirical strategy. In Section 4.4 we establish the link between our proxy for labor market slack introduced in equation (13), and inflation. We present the main results of the paper in Section 4.5. Section 4.6 presents micro evidence on the behavior of the on-the-job search rate. Finally, we discuss the performance of the model in fitting inflation on a longer sample starting in the early 1990s in Section 4.7.

## 4.1 Calibration

We calibrate the steady state of the model to the US 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 eleven parameter values: $\beta$, $\phi_0$, $\delta$, $y_b$, $y_g$, $\upsilon$, $b$, $\xi_g$, $c$, $c^f$ and $s$. We set the discount factor $\beta$ to 0.9987505, in order to match an annual real interest rate of 1.5%, which is in line with the median of individual economic projections about the real long-term interest rate from various Federal Reserve's Board members, FOMC members, or FOMC participants (known as *Survey of Economic Projections*, SEP).[11] We normalize $\theta$ to unity, which allows us to pin down the scale factor $\phi_0$, so as to match a job finding rate of 33 percent, which is the average of the job finding rate computed

[11] We take the average of these projections 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 |
| *Parameters that do not affect the steady-state* | | | |
| $\chi$ | Mark-up parameter | 1.2000 | 20% mark-up |
| $\varkappa$ | 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 | MLE estimation |
| $100\sigma_S$ | St. dev. of job search rate shocks | 2.5510 | MLE 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) |
| $EE \equiv \frac{\frac{1}{\upsilon}s\phi\left[l_b^0\left(\xi_b+\upsilon^{-1}\xi_g\right)+l_g^0\xi_g\right]}{l_b^0+l_g^0}$ | EE transition rate | 0.0258 | 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 top10% 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.

following Shimer (2005) over the last 25 years (January 1993-December 2018).[12] The job separation rate $\delta$ is implied by the Beveridge curve, under the assumption of a steady state rate of unemployment of 5.5%. Namely, solving the Beveridge curve for $\delta = \frac{\phi_0 u_0}{1-u_0+\phi_0 u_0}$ yields a separation rate of 0.02. The productivity of a bad match is normalized to one and the productivity in a good match is set to be 8% higher. We regard this productivity differential as conservative, in the light of values that have been assigned in the calibration of other comparable models with on-the-job search. Our targeted wage differential is in line with evidence by Faberman et al. 2019 based on the *Survey of Consumer Expectations*, who show that wage gains associated with job switching are about 8%, after controlling for observable characteristics of workers and jobs.

[12]Under the assumption of unitary tightness ($\theta = 1$), the job finding rate becomes equals to $\phi_0$.

Moreover, we noticed that assigning 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. Finally, we set the probability that workers will accept an equally valuable outside offer to be $\upsilon = 0.5$. This value is large enough to allow the model to match the average EE flow rate in the U.S. economy. In Appendix G, 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: *(i)* A value of expected hiring costs, including both the variable and the fixed cost component, equal to two weeks of wages.[13] *(ii)* A fraction of good jobs in steady state equal to 67%, which is the share of employment for the top 10% US firms by employment size in 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 US, following the analysis of Silva and Toledo (2009), which is based on the 1982 Employer Opportunity Pilot Project (EOPP), a cross-sectional firm-level survey that contains detailed information on both pre-match and post-match labor turnover costs in the United States.[14] *(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 (April 1990 through December 2007). 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). We have checked that the value of $b$ implied by the calibration is consistent with a positive surplus for low quality matches both in steady state and in every month considered in the empirical exercise of Section 4.5.

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

Turning now to the parameters that do not affect the steady-state, we set the smoothing coefficient of the Taylor rule to the value of 0.85, which corresponds to a coefficient of around

[13] The average wage is measured as the price of the labor service $\varphi$.

[14] 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\$.

0.65 in quarterly space, and the response parameters to inflation and output to the values of 1.8 and 0.25, respectively. The parameter $\chi$ is set to equal 1.2, which implies a 20% price mark-up. The Calvo parameter governing price stickiness is set to 0.925, which in quarterly frequency implies a probability of not readjusting prices equal to 0.8. The scale parameter of the Kimball aggregator is set to 10, the value used by Smets and Wouters (2007). This value for the Kimball parameter lowers the sensitivity of inflation to real marginal costs, allowing us to get a plausible volatility of price inflation. The steady state gross rate of inflation is set to equal 1.0016, which implies a 2% annualized rate of inflation. Finally, we set the elasticity of vacancies in the matching function $\psi$ to equal 0.5 to be consistent with estimates by Moscarini and Postel-Vinay (2018), which account for workers searching on the job.

As we will show in Section 4.3, we can use the observed unemployment rate and the EE flow rate in combination with a subset of model equations to obtain the 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.

Turning to the parameters affecting the persistence and the volatility of the preference shock, we set the autocorrelation parameter, $\rho_\lambda$, to 0.80 and then we calibrate the standard deviation, $\sigma_\lambda$, so that the model can match the volatility of the observed unemployment rate in the data (January 1992 - December 2018).[15] 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 to propagate as a supply shock moving unemployment rate and inflation in the same direction.[16] Since the other shock (i.e., 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 commove.

**Model Solution with the Zero Lower Bound (ZLB) Constraint** The model is log-linearized around its steady state equilibrium.[17] However, the zero lower bound introduces a nonlinearity that prevents us from solving the model with standard solution methods. We develop a novel method to find the certainty-equivalence solution to these temporarily non-linear dynamics. Our method does not require us to assume that agents in the model have

[15] We pick the unemployment rate as a target variable because it will be used in our main empirical exercise.

[16] 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 $\varphi_t$. In turn, the rise in current and future expected marginal costs entails a rise in the rate of inflation, together with a contraction in aggregate production.

[17] Rates and shares are linearized, all the other variables are loglinearized.

perfect foresight. Agents update their rational expectations about the duration of the zero lower bound over time after having observed past and current shocks.

Our method relies on appending a sequence of anticipated shocks (dummy shocks) to the unconstrained Taylor rule. Anticipated shocks are known by agents in the current period but will hit the economy in future periods. The sequence of these shocks is computed so as to ensure that agents expect that the zero lower bound constraint will be satisfied for the next 36 months in every period.[18] When the constraint is not expected to become binding, these anticipated shocks are set to zero. Obviously, these shocks will have an effect on the expected duration of the ZLB and hence on equilibrium outcomes, requiring us to solve a fixed point problem, which is described in greater detail in Appendix H. This fixed-point problem does not turn to be time consuming or computationally challenging in practice.

## 4.2 Impulse Responses

In this section we discuss the propagation of the two shocks of the model: the preference shock and the shock to the rate of on-the-job search. We start with the latter shock, whose propagation has been informally discussed in the partial-equilibrium thought experiment of Section 3.6. Figure 3 shows that a fall in the rate at which workers search on the job raises the fraction of job seekers that are unemployed (i.e., the first component on the right hand side of equation (13) defining labor market slack in our model), lowering the intensity of wage competition and increasing slack. In expectation, producing labor service becomes cheaper for an entrant firm as the likelihood of extracting a positive surplus from meeting a worker increases. In addition, the stock of bad matches rises and the stock of good matches drops. This is quite mechanical as this shock directly reduces the flow from bad to good jobs, slowing down the allocative mechanism of the ladder. This increase in labor misallocation implies that wage competition is less likely to entirely wipe out the firms' share of surplus and, as a result, the second component of the right-hand-side of equation (13) rises, implying a further decline in the intensity of wage competition among firms and a further increase in slack. As the likelihood of being engaged in a wage competition that will zero the surplus for entrant labor service firms falls, inflation drops and the central bank cuts the interest rate, stimulating aggregate demand and reducing unemployment. Moreover, attracted by the expectation of cheaper labor, more firms enter the labor service sector, i.e. more vacancies are created, expanding aggregate supply, which also contributes to lowering the unemployment rate.

Note that the fall in the unemployment rate, in isolation, contributes to lower the probability for an entrant firm to meet an unemployed worker and hence causes wage competition to become

[18] In none of the periods of our sample, the zero lower bound constraint binds for more than 36 months in expectation. If it did, we would need to add more anticipated shocks to the Taylor rule so as to cover a horizon longer than 36 months.

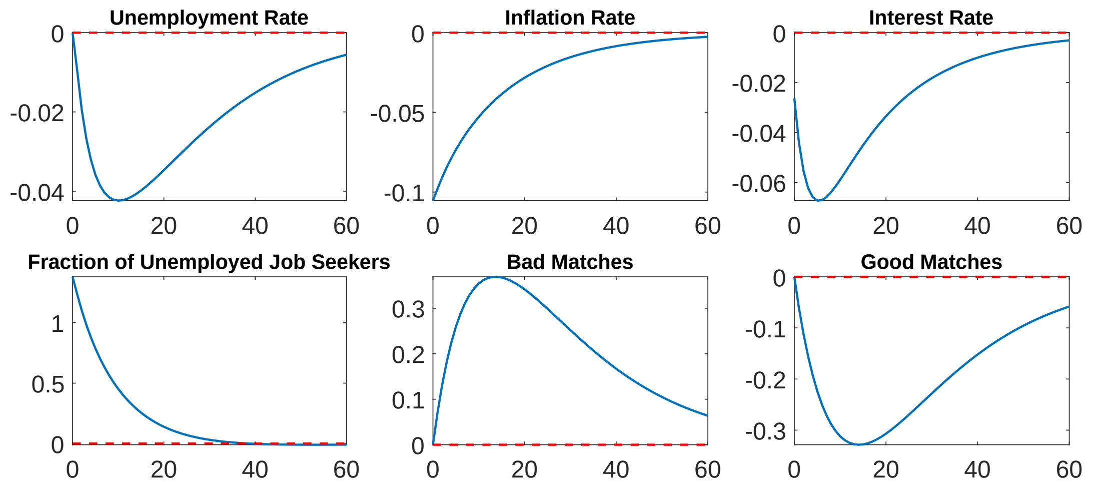


Figure 3: Impulse responses to a shock that lowers the on-the-job search rate by one standard deviation. Units: percentage points. Inflation and interest rate are expressed in annualized rates.

more intense. Yet, as shown in the lower left graph of Figure 3, it turns out that in equilibrium 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.

By showing the response of the fraction of job seekers who are unemployed and that of the stock of workers employed in bad matches, we want to provide a decomposition of our measure of slack defined in equation (13). While in the immediate aftermath of the shock, inflation responds mostly to the rise in the fraction of unemployed job seekers, the persistent change in the match composition of the employment pool weights down on inflation later on, contributing to keeping inflation below its long-run value for some time. 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 US economy in the most recent years.

Figure 4 shows the responses to a negative preference shock. As before, 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 rate to fall. Concurrently, the weakening of the price setters' demand for labor services 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, labor becomes cheaper in expectation for an entrant

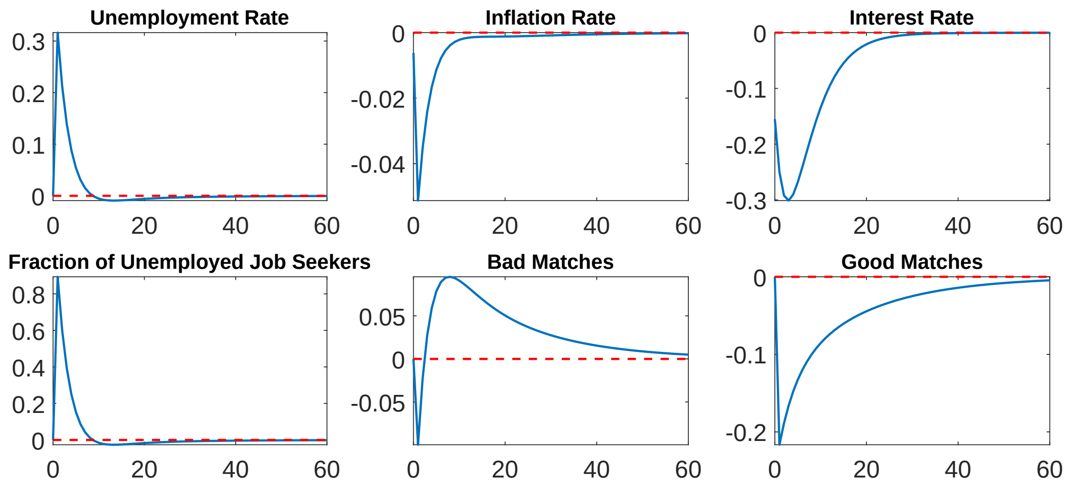


Figure 4: Impulse responses to a negative preference shock by one standard deviation. Units: percentage points. Inflation and interest rate are expressed in annualized rates.

service firm, since it is now more likely to extract a nonzero surplus from the match. As a result, equation (13) implies that labor market slack increases, the price of the labor service falls, and hence inflation drops even further 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. Our measure of slack captures these effects through the second component in the right-hand-side of equation (13).

In analogy with the case of the shock to the rate of on-the-job search, in the immediate aftermath of a preference shock the dynamics of inflation reflect mostly the response of the fraction of unemployed job seekers. But as the on-the-job search rate converges to its steady-state value, the effects of labor misallocation on labor market slack takes over, raising the persistence of inflation after the shock.

## 4.3 Measuring Labor Market Aggregates

It is important to notice that for a given value of bad and good matches at the beginning of the sample, observing the unemployment rate and the EE rate implies the entire time series of the on-the-job search rate, $s_t$, as well as the time series of bad and good matches, $l^0_{b,t+1}$ and $l^0_{g,t+1}$. The exact identification of these variables comes from a set of identities and does not require solving the model.

We first show this property of the model. Then we use the observed series of the unemployment rate and the EE rate to actually recover the on-the-job search rate and the share of bad matches. We use the equations linearized around the steady-state equilibrium where $\tilde{}$ denotes linearized variables.

The observed series of unemployment rates informs $u_{0,t+1}$ and hence the aggregate unemployment at the end of the period, $u_t$, through the following equation

$$\tilde{u}_t = \frac{\tilde{u}_{0,t+1}}{1-\delta}, \tag{19}$$

which is is obtained by combining equations (3) and (4) and linearizing.

Endowed with the end of period unemployment rate $\tilde{u}_t$, we can linearize equation (5) to pin down the job finding rate $\tilde{\phi}_t$ at time $t$ as follows:

$$\tilde{\phi}_t = \frac{(1-\phi)\,\tilde{u}_{0,t} - \tilde{u}_t}{u_0}, \tag{20}$$

where $u_0$ denotes the unemployment rate at the beginning of the period in steady state and $\phi$ is the job finding rate in steady state. We iterate on equations (19) and (20) using the observed series of the unemployment rate, which yields a time series for the job finding rate $\tilde{\phi}_t$.

We then linearize the definition of the EE flow rate, $EE_t$, in the model, which reads:

$$EE_t \equiv \frac{\upsilon s_t \phi(\theta_t)\left[l_{b,t}^0\left(\xi_b + \upsilon^{-1}\xi_g\right) + l_{g,t}^0 \xi_g\right]}{l_{b,t}^0 + l_{g,t}^0}. \tag{21}$$

The EE rate is the ratio of how many workers employed at the beginning of the period have switched job (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 find an equally valuable match.

The linearized equation defining the EE rate above reads as follows:

$$\begin{aligned}\tilde{s}_t \quad = \quad & \frac{s}{EE}\widetilde{EE}_t - \frac{s}{\phi}\tilde{\phi}_t - \frac{s}{\upsilon\left(l_b^0 + l_g^0\right)}\left[\frac{s\phi\left[\left(\xi_b + \upsilon\xi_g\right)\right]}{EE} - 1\right]\tilde{l}_{b,t}^0 \\ & - \frac{s}{\upsilon\left(l_b^0 + l_g^0\right)}\left[\frac{s\phi\xi_g}{EE} - 1\right]\tilde{l}_{g,t}^0. \end{aligned} \tag{22}$$

Since $\tilde{l}_{b,t}^0$ and $\tilde{l}_{g,t}^0$ are predetermined at time $t$, this equation allows us to exactly measure the on-the-job search rate $\tilde{s}_t$ consistently with the series for the job finding rate $\tilde{\phi}_t$ and the observed EE flow rate $\widetilde{EE}_t$.

With the rates $\tilde{\phi}_t$ and $\tilde{s}_t$ at hand, we can use the observed unemployment rate $\tilde{u}_{0,t}$ to 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 (16) and (17), which read

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

$$\tilde{l}^0_{g,t+1} = (1-\delta)\left[\tilde{l}^0_{g,t} + \phi\xi_g l^0_b \tilde{s}_t + s\phi\xi_g \tilde{l}^0_{b,t} + \phi\xi_g \tilde{u}_{0,t} + \left[s\xi_g l^0_b + \xi_g u_0\right]\tilde{\phi}_t\right]. \tag{24}$$

With the knowledge of the distribution of match quality at time $t+1$, we can go back to equation (22) and obtain the on-the-job search rate in period $t+1$ ($\tilde{s}_{t+1}$). Repeating these steps will give us a series for the on-the-job-search rate and for the distribution of match quality in our sample. Note that this procedure allows us to also obtain the series of labor market slack by using equation (13).

It is important to notice that solving the model is not needed to pin down exactly the series of the on-the-job search rate. This property of the model allows us to estimate the parameters $\rho_s$ and $\sigma_s$ before solving the model. This procedure is conditioned on the fraction of bad matches at the beginning of the sample period (in our case April 1990). We assume that the distribution of match quality is at steady state at that point in time.[19]

One concern with this approach is that by relying entirely on the unemployment rate to estimate the job finding rate in equations (19)-(20), we are not taking into account changes in the separation rate or in the participation rate, potentially leading to biased estimate of the job finding rate. This is a concern because this bias could distort our estimate of the on-the-job search rate and of labor misallocation. A way to mitigate this problem is to directly use the job finding rate measured in the data in place of the observed rate of unemployment and obtain the series of the on-the-job search rate and those of the good and bad matches by iterating on equations (22) - (24). At the same time though, using job finding rates would bias the implied unemployment rate, which is one of the three key drivers of inflation in the model. Our main results are not affected by using the job finding rate computed following Shimer (2005) instead of the unemployment rate as an observable. The only noticeable difference is in the estimated behavior of labor market slack over the Great Recession period, when the separation rate spiked up in the data in a way that is not captured by the model. Since this effect is moderate and contained within a handful of quarters, it does not materially affect our analysis on the pre-Great Recession data, which will be shown in Section 4.7.

**On-the-Job Search Rate and Bad Jobs in the Data** We use two monthly time series to measure the on-the-job search rate and the other labor market variables. The first series

[19] Results would not change if we introduce a Gaussian prior reflecting uncertainty about the initial conditions and then use the Kalman filter to optimally estimate these initial conditions.

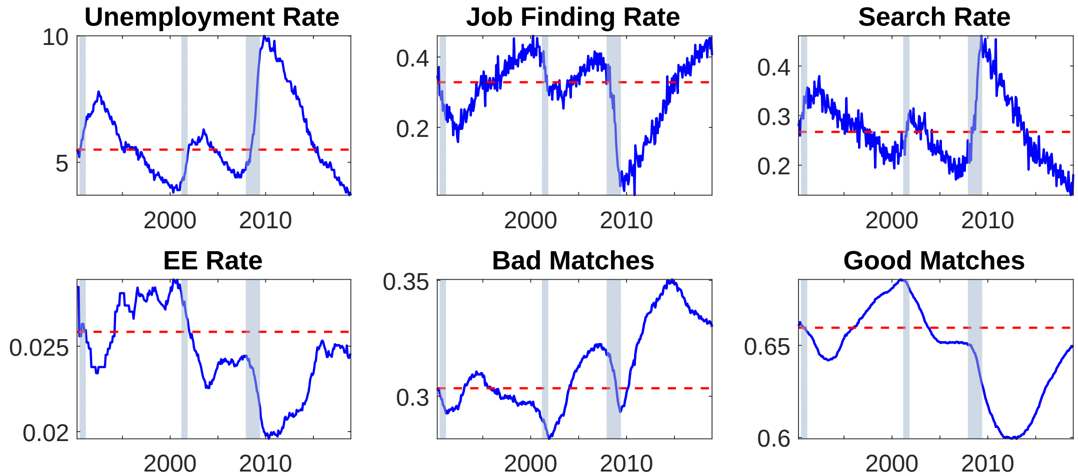


Figure 5: Labor market variables simulated from the calibrated model using the shocks that allow the model to explain exactly the observed unemployment rate and the EE flow rate, which are shown in the left graphs. The red dashed lines denote the model's steady state value of the simulated variables.

is the civilian unemployment rate. The second series is the EE flow rate measured from the CPS data by Fujita, Moscarini, and Postel-Vinay (2019) and extended back to April 1990 by splicing this series with the quit rate measured by Davis, Faberman, and Haltiwanger (2012).[20] 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. We think that this is interesting given that, to our knowledge, this is the first paper that provides their estimation using aggregate labor market flows.

Figure 5 shows the dynamics of the rate of on-the-job search, $\tilde{s}_t$, and bad jobs, $\tilde{l}^0_{b,t}$, along with the two traditional labor market variables -the unemployment rate and the job finding rate rate $\tilde{\phi}_t$ - over the sample period that goes from April 1990 through December 2018. The panels on the left report the observable variables. While the traditional measures of labor market slack, such as the unemployment and the job finding rate, reported in the upper panel of Figure 5, suggest that the U.S. labor market has become quite tight in recent years, the dynamics of two key drivers of the model's labor market slack in equation (13), i.e. the on-the job-search rate and the stock of bad matches, paint a different picture: since the end of the last recession, the rate of on-the-job search has fallen to a historically low level, and bad matches have increased, remaining at a high level throughout the recovery. This latter finding implies that cyclical misallocation is still high, bearing down on inflation and labor productivity.

Quite interestingly, while the amount of good matches has been chugging along well in

[20]While the quit rate also includes workers who leave their jobs to become unemployed or to exit the labor force, Elsby, Hobijn, and Sahin (2010) document that 86% of the workers observed quitting their jobs move directly to a new job.

recent years and is now close to its long-run value, the convergence of bad matches has slowed down markedly. This pattern suggests that the low unemployment rate has led to the creation of a large mass of low-productivity jobs that will be converted to high-productivity jobs only slowly because of the record low rate of on-the-job search.

The prediction that bad jobs are still heightened 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.[21] This increase in bad jobs also accords well with the findings in Jaimovich et al. (2020), who show that a third of the workers who were employed in routinary occupations before the Great Recession could not find similar jobs and are now stuck in nonroutinary manual occupations.

Looking at the longer sample, the on-the-job search rate implied by the model and reported in the top right corner of Figure 5 exhibits a clearly countercyclical pattern. In recessions, the job finding rate falls more than the EE rate. If the job finding rate captures the arrival flow of job opportunities that applies equally to both workers searching on and off the job, then it must be that in recessions employed workers search more.[22] In expansions instead, the job finding rate recovers more quickly than the EE rate. If job opportunities increase equally for both types of workers, then it must be that the rate of on-the-job search falls. To put it differently, the countercyclicality of the on-the-job search rate is due to the higher volatility of the job finding rate relative to the EE rate. This observation implies that the dynamics of the rate of on-the-job search are mainly driven by the job finding rate. Since the job finding rate enters with a minus sign in equation (22) and is a strongly procyclical variable, the on-the-job search rate has to be countercyclical.[23]

A number of explanations could support this countercyclical behavior of the on-the-job search rate. The decision to look for jobs is likely to be positively related to individual income risk, which is countercyclical. So it may be that on average, fewer employed workers search in expansions simply because less of them feel at risk of losing their jobs. To the extent that this behavior dampens the volatility of the EE rate over the business cycle, our model rationalizes it with a countercyclical rate of on-the-job search. But the countercyclicality of on-the-job search may as well derive from compositional effects, which could also affect the dynamics of the EE flow rate in the data. Workers may search harder and hence switch jobs more often when they are employed in bad matches, which is prevalent at the beginning of an expansion. This view is consistent with the findings in Faberman et al. (2019), who show that employed workers search

[21] The question asks: "On a scale from 1 to 7, how well do you think this job fits your experience and skills? About 30% of the respondents report a satisfaction of 5 or less.

[22] An alternative explanation is that the probability of forming a good match rises in recessions, i.e. $\xi_g$ rises. This possibility runs counter to the empirical evidence reviewed by Barlevy (2002).

[23] Kudlyak and Faberman (2019) observe the job application behavior of the users of Snag-A-Job, an online job site, and find results that are consistent with the search intensity of the employed being countercyclical.

more intensively, the lower their residualized wage. In addition, workers who are hired at the beginning of an expansion are generally more skilled and dynamic than those who tend to find jobs when the labor market is already very tight. This view is consistent with the findings in Cahuc, Postel-Vinay, and Robin (2006), who show that workers in higher skill categories tend to be more mobile than lower skilled ones. To the extent that these mechanisms influence the behavior of the EE flow rate significantly, the model will predict the rate of on-the-job search to be countercyclical.

Taken together, all these explanations also suggest a possible reason why the search rate in Figure 5 has fallen to its historical through at the end of the sample. The very prolonged fall in the rate of on-the-job search might be related to the exceptionally long expansion the U.S. economy is going through. Another interesting finding is that bad matches fell in all the three recessions in our sample. This decline is due both to the prevalence of negative preference shocks, which reduces vacancies and hence the inflow of workers back onto the ladder, and to the increase in the countercyclical rate of on-the-job search, which raises the speed at which employed workers get reallocated to better jobs.

## 4.4 A Useful Decomposition

The behavior of inflation in the model closely reflects the contemporaneous probability that, conditional on a contact, labor service firms are not engaged in a wage competition that leads them to relinquish the entire surplus to the worker. This probability is defined in equation (13). In Section 3.6, we explained that an increase in this probability is tantamount to a rise in labor market slack, which reduces the expected costs of the labor service and hence induces price setters to lower their price. Indeed, when the model is simulated for a large number of periods, the correlation between slack and the month-over-month inflation rate is $-0.94$.

Equation (13) also makes it clear that our measure of slack is determined by three components: the unemployment rate, $u_{0,t}$, the measure of bad matches, $l_{b,t}^{0}$, capturing labor misallocation, and the share of workers searching on the job, $s_t$. We conjecture that a linear combination of these three components is also key to determine the contemporaneous rate of inflation in the loglinearized model. This conjecture stems from our discussion in Section 3.6 and the inspection of the impulse response functions in the previous section. To verify our conjecture, we simulate the calibrated model for a large number of periods (one million) and then regress the simulated series of inflation on the three determinants of slack defined in equation (13). This procedure gives us three weights that maximize the explanatory power of the three drivers of labor market slack on inflation. The weights are as follows:

$$\hat{\pi}_t = \underset{[-0.9722,\quad -0.9716]}{-0.9719}\ \tilde{u}_{0,t} - \underset{[-0.4519,\quad -0.4515]}{0.4517}\ \tilde{l}_{b,t}^{0} + \underset{[0.2538,\quad 0.2540]}{0.2539}\ \tilde{s}_t, \tag{25}$$

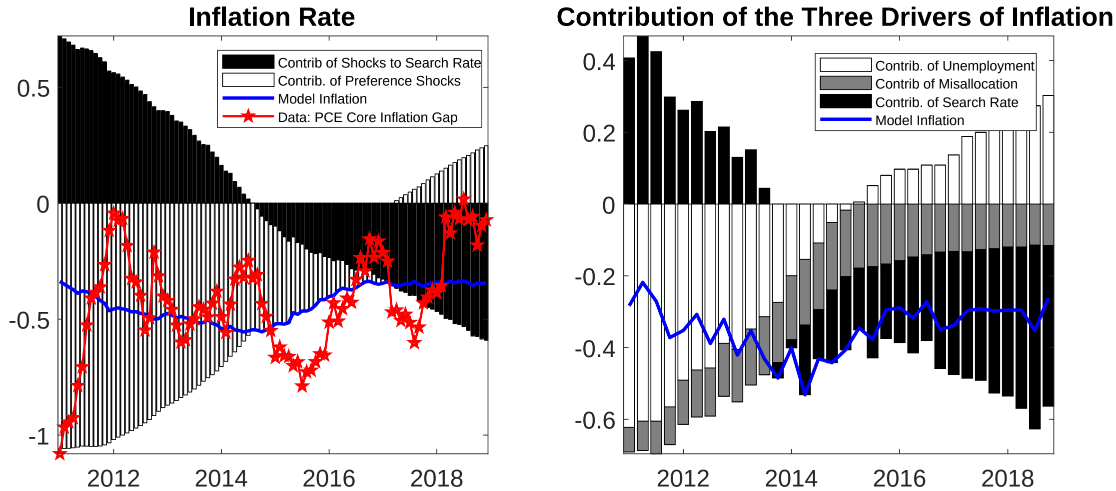


Figure 6: Left Graph: PCE core inflation gap in the data (red line with the star markers) and model's corresponding inflation rate in deviation from its steady state value (blue solid line). The bars denote the contribution of the search rate shocks (black bars) and that of the preference shocks (white bars) to model's inflation. The core PCE inflation gap is obtained by subtracting the ten-year PCE inflation expectations measured by the Survey of Professional Forecasters from the year-over-year core PCE inflation rate. Model's inflation is also computed as the year-over-year inflation rate. All rates are in percent and annualized. Right graph: Annualized quarter-over-quarter inflation rate (blue solid line) and its decomposition into its three main drives (bars). We use the estimated equation (25) to decompose inflation.

where we report the 95-percent confidence interval for the coefficient within square brackets under the estimated value.

The R-squared of the OLS regression is 0.9922 to signify a close-to-perfect ability of the three labor market variables to explain contemporaneous inflation in the model, which confirms our initial conjecture. While the three components in the right hand side of equation (25) are not derived from a formal notion of output gap in the model, in practice, as we shall show, they allow for a decomposition of inflation that turns out to be very useful in interpreting the results of the paper in the next section.

## 4.5 The Missing Inflation in the Post-Great Recession Period

We want to evaluate the ability of the model to explain the missing inflation during the recovery that followed the Great Recession. We are particularly interested in this period since the conventional theories of inflation more clearly fail, as shown in Section 2. The data set is identical to the one used to measure the on-the-job search rate in Section 4.3. We use our loglinearized model to retrieve the series of the two shocks that make the model explain exactly the observed unemployment rate and EE flow rate. We then feed the model with these shocks to simulate the inflation rate predicted by the model in the last decade.[24]

The left graph in Figure 6 illustrates the main results of this exercise by comparing the

[24] 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 since the beginning of the sample is in April 1990 and the analysis focuses on a sample period that starts several years later; specifically January 2011.

inflation rate in the data to the rate of inflation simulated from the model, and its shock decomposition. The red starred line denotes the observed core PCE inflation gap, which is obtained by subtracting the ten-year PCE inflation expectations measured by the *Survey of Professional Forecasters* from the year-over-year core PCE inflation rate. The solid line denotes the corresponding measure of inflation predicted by the model, using the simulation procedure described earlier.[25] The black and white bars indicate the contributions of the shocks to the search rate and to preferences, respectively. The bars should be interpreted as the inflation rate predicted by the model when we feed it with each one of these shocks.

The model can explain why inflation has remained persistently below the long term expectations in the last decade. As illustrated by the black bars, the model attributes the missing inflation over the last decade to the decline in the rate of on-the-job search. This fall has reduced the intensity of wage competition for employed workers throughout the recovery, generating a fair amount of deflationary pressures, in spite of the steady decline in the unemployment rate. The preference shocks, which are identified by the rate of unemployment, capture the state of the business cycle and the effects of the ZLB constraint. The white bars in the left graph of Figure 6 show that these factors contribute to generating deflationary pressures in the immediate aftermath of the crisis and positive inflationary pressures over the most recent years. Nevertheless, the deflationary pressures due to the fall in the rate of on-the-job search (the black bars) more than compensate for the inflationary pressures due to the preference shocks (the white bars) in recent years. In accordance with the impulse responses shown in Section 4.2, the fall in the rate of on-the-job search contributes to increasing production and to lower the rate of unemployment, while exerting downward pressure on the rate of inflation.

We now use the decomposition of inflation introduced in Section 4.4 to provide further intuition about which factors are contributing to the missing inflation. The right graph of Figure 6 visualizes the decomposition of model's inflation into its three main drivers: the unemployment rate, the stock of bad jobs, and the on-the-job search rate. In the graph, model inflation is expressed at quarterly frequencies by taking the average of monthly figures.[26] At the beginning of the recovery, inflation has been low primarily because of the record surge in the unemployment rate during the Great Recession, as illustrated by the white bars. After 2015, 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, in the same years, the on-the-job search rate declined rapidly, putting downward

[25] The PCE inflation rate expected by the professional forecasters ten year from now is extremely stable around two percent. The series is shown in the right plot of Figure 1.

[26] This is one reason why model inflation is not exactly identical in both graphs. Another reason is that the inflation rate shown on the left graph is the year-over-year growth rate of the price level, to make it comparable with how we constructed the PCE core inflation rate (the starred red line in the left graph) whereas inflation in the right plot is the quarter-over-quarter inflation rate. Notwithstanding these two caveats, the model inflation rates shown in the two graphs are very similar.

pressures on inflation (the black bars) and dominating the effects of the unemployment rate (the white bars).

The role played by the incidence of bad matches is also very interesting (the gray bars in the right graph of Figure 6). Bad matches have always contributed to keeping inflation below its long-run level. In the earlier part of the period of interest, the model predicts that following the unusually severe recession, a large fraction of unemployed workers took a first step onto the ladder, raising the stock of bad jobs. This pattern is consistent with the propagation of preference shocks to the share of bad matches shown in Figure 4 and is fairly typical in this class of models as it takes time, after a worker loses her job, to climb the ladder all the way up again. 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 clearly highlight the important role played by the cyclical match composition of the employment pool in explaining the missing inflation.

**What is So Special About the Great Recession?** Figure 5 suggests that there may be nothing special about the Great Recession, in terms of its long-lasting implications for inflation, since the countercyclicality of the on-the-job search rate emerges as a striking empirical regularity. In putting upward pressure on inflation in recessions and downward pressure on inflation as economic recoveries progress, such a countercyclicality of the rate is consistent with the lack of serious deflationary and inflationary episodes in the U.S. over the last twenty years. We mainly focus on the recent recovery because this is a period in which the traditional theories of labor market slack more spectacularly fail to explain inflation, as shown in Section 2.

**The Role of Labor Costs** It has been argued that the missing inflation is mainly due to the decoupling between the dynamics of labor costs and inflation (Belz, Wessel and Yellen, 2020). The argument is that the tight labor market has boosted wage growth but not inflation. According to this view, the inflation Phillips curve no longer seems to be borne out by the data whereas the wage Phillips curve, which links labor market slack to wage inflation, fares substantially better empirically. We do not see strong empirical support to this view at least in the last decade. As a counterexample, if one looks at the dynamics of average hourly earnings in the last decade, there is not strong evidence of wage pressures even though the traditional measures of slack suggest that the labor market is very tight.

The left graph of Figure 7 compares the 12-month moving average of the month-over-month growth rate of the nominal marginal costs simulated from the calibrated model (blue solid line) to the 12-month moving average of the month-over-month of the U.S. average hourly earnings (black dashed line). The horizontal red dashed line denotes the average wage growth in the data. The mean and the volatility of the model's implied growth rate of marginal costs are

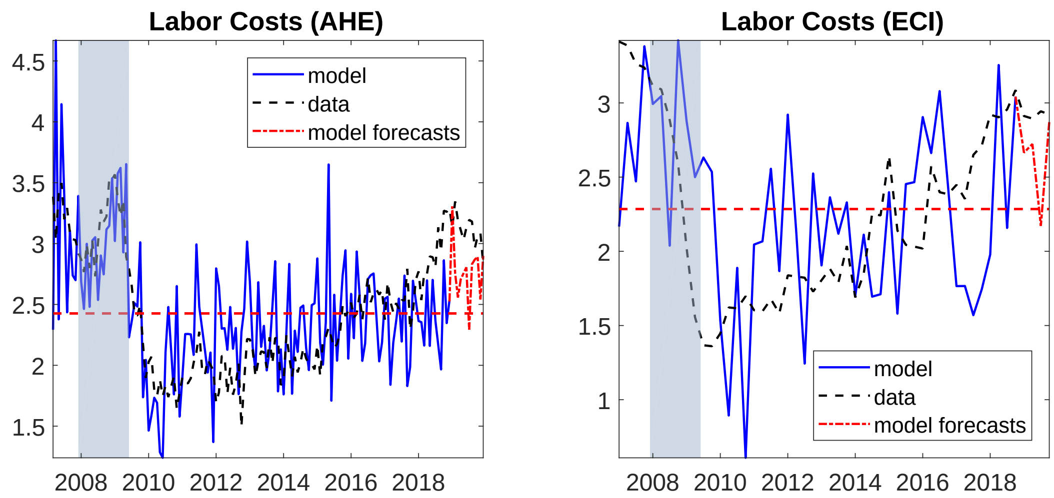


Figure 7: Left Graph: Nominal marginal costs growth rate in the model and the growth rate in the average hourly earnings of all employees: total private (CES0500000003). Moving average over the last 12 months. Source: BLS. Units: Percentage points of annualized rates. Frequency: monthly. Right Graph: Nominal marginal costs growth rate in the model and the growth rate in the Employment Cost Index: Wages and Salaries: Private Industry Workers (ECIWAG). Moving average over the last four quarters. The series of nominal marginal costs implied by the model are made quarterly by taking the average of monthly growth rates of the nominal marginal costs. Source: BLS. Units: Percentage points of annualized rates. Frequency: quarterly. In both graphs, the red-dashed line marks the model's forecasts of nominal marginal costs growth.

rescaled to match those in the data. Rescaling the volatility of the model's series facilitates the readability of the graph and is neutral with respect to the in-sample correlation of the two plotted series.[27]

Two important lessons emerge from this graph. First, we do not see an acceleration in wage growth in recent years, in contrast to what would be implied by a standard wage Phillips curve combined with the various measures of slack plotted in Figure 2. Second, our model explains the lack of upward pressures on labor costs observed in the data with the decline in the on-the-job search rate, which has brought about a persistent fall in the intensity of wage competition. Similar conclusions can be reached by looking at the graph on the right of Figure 7, which compares the model's predicted growth rate of nominal marginal costs with the growth rate in the Employment Cost Index measured by the BEA.

Since our data set ends in 2018Q4, we show the model's forecasts for the last twelve months of data (the red dashed line). We include the 2019 data on wage growth, showing that it has stopped increasing. This fact is arguably difficult for standard models to explain given the rapid increase in their measures of slack shown in Figure 2. While our model sees wage growth close to its long-run value (red dashed line), it does not predict significant wage pressures in 2019.

[27]The assumed bargaining protocol, which was introduced in Section 3.5, leads to excess volatility of marginal costs. This volatility could be reduced by introducing ad-hoc assumptions (e.g., wage rigidities and indexation). Since the main focus of the paper is on the missing inflation not on wage dynamics, we do not make these assumptions that would complicate the already involved economics of the model and would obfuscate intuition. The assumption of Kimball aggregator is precisely introduced to ensure a plausible volatility of inflation.

**Endogenizing the On-the-Job Search Rate** In this paper we treat the rate of on-the-job search as an exogenous process. We believe that this is the right approach at this stage since the empirical micro labor literature has not yet settled on what could be a plausible theory of what drives this rate. It is important to notice that the exact identification of the rate of on-the-job search implies that endogenizing the search decision would not affect the in-sample estimation of this rate, which is shown in the upper right graph of Figure 5. It follows that any way of microfounding the households' decision about the on-the-job search would matter for our analysis only to the extent that it alters agents' expectations about the future evolution of the rate.

In the model, agents form expectations about the likely evolution of the on-the-job search rate by using an AR(1) process whose parameters ($\rho_S$ and $\sigma_S$) are estimated via maximum likelihood, as explained in Section 4.1.[28] In Appendix G, we show that reasonable deviations from the maximum likelihood estimates of the autocorrelation parameter, which could be warranted by the microfoundation of the dynamics of the on-the-job search rate, do not materially affect the model's predicted path for inflation in the past decade.[29] This result suggests that endogenizing the decision to search on the job would not significantly alter the conclusions of the paper about the model's ability to explain the missing inflation.

## 4.6 The On-the-Job Search Rate in the Micro Data

In the previous section we have illustrated that, according to our model, the main reason why inflation has remained below target even in the most recent years is because of the steady fall in the rate of on-the-job search, $s_t$. As explained in Section 4.3, given the assumptions of the model, $s_t$ is implied by the time series of EE and unemployment rates.

In this section, we look into the micro data to see if these findings are validated at the micro level. To this end, we explore a new survey that is informative of the search behavior of the employed workers, and that 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

[28] The AR(1) process is the best at fitting the time series of the on-the-job search rate.

[29] The most critical case is when the on-the-job search rate follows an i.i.d. process. This assumption causes expectations about the rate to revert back to steady state in the next month. This quick mean reversion implies that agents expect a drastic rise in the interfirm wage competition in the next period. These beliefs could raise model's prediction of inflation today in our forward-looking model. Hence, this is the case that could potentially undermine the model's ability to explain the missing inflation. However, we find that the model's in-sample predictions about inflation vary only marginally when the search rate is assumed to be i.i.d as shown in Appendix G. Also note that we do not perturb the standard deviation of the shocks to the on-the-job search rate, $\sigma_S$, as doing so would quite clearly have no effect whatsoever on expectations and hence on the model's predicted path of inflation.

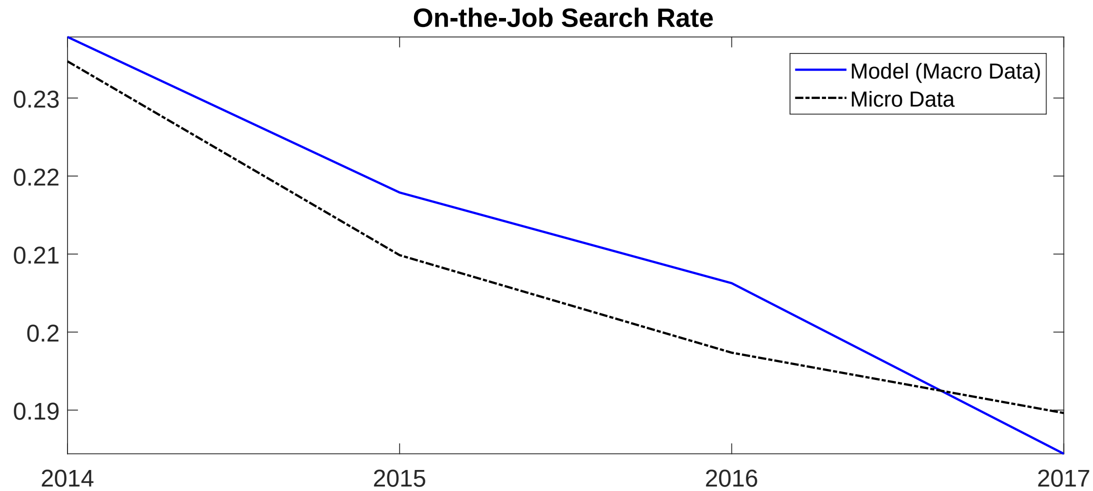


Figure 8: The on-the-job search rate in the model and in the *Survey of Consumer Expectations*. In the *Survey*, the rate is computed by dividing the workers who have searched at least one hour within the last seven days by the total number of workers surveyed. The rate is conditional on those surveyees who are working for someone else.

for work in the previous seven days.[30] Because it is a recent survey, data are only available from 2014 to 2017. Even if this is admittedly a very limited period of time, it still covers four years in which our model predicts that the on-the-job search rate was below its long-run value and kept falling down.

Figure 8 plots the on-the-job search rate implied by the model, $s_t$, and the corresponding measure in the microdata (blue solid line and the dashed-dotted black 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 microdata.

When the model's variable $s_t$ is measured from equation (22), 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 more stringent hiring practices.

Using information on the hours of search for the employed workers in SCE, we find that the fall in the aggregate amount of time spent searching is entirely explained by the extensive

[30]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.

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 have also looked at how the arrival rate of job offers for the employed workers has varied over the time relative to the arrival rate of offers for the unemployed. That is, we have 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. The ratio of these two ratios does not exhibit a clear pattern. Therefore, this *Survey* validates the decline of the on-the-job search rate predicted by our macro model, at least for the period in which micro data are available.

But why has the rate of on-the-job search declined over the post Great-Recession recovery? Investigating the determinants of the search behavior of the employed workers, an avenue of empirical labor research that has recently been revived by the availability of new sources of information,[31] is outside the scope of this paper. However, we discuss a number of potential explanations that are consistent with the pattern that we observe. In doing so, we link our investigation to other developments in the literature.

One possibility, as discussed in Section 4.3, is that the rate of on-the-job search is particularly low simply because it is countercyclical and the expansion has been going on for an extraordinarily prolonged period of time. Alternatively, the decline in this rate in recent years can be explained by the discouragement of those employed workers who have experienced the disappearance of suitable jobs. Jaimovich et al. (2020) show that a third of the workers that were employed in routinary occupations before the Great Recession could not find similar jobs and are now stuck in nonroutinary manual occupations. Furthermore, Jaimovich and Siu (2018) show that the incidence of job polarization is higher in recession. To the extent that learning about one's grimmer employment prospects takes time and may lead to discouragement, job polarization is consistent with the fall in the on-the-job search rate in recoveries.

Job polarization has also implications for the low-frequency behavior of the on-the-job search rate. While we find that the on-the-job search rate has reached a record low in recent years, we could not find any conclusive evidence to prove or disprove the existence of an active trend in the behavior of the on-the-job search rate in the last thirty years, which is shown in the upper right graph of Figure 5. Therefore, we did not introduce a trend in the process driving the search rate. It can be shown that adding such a trend would greatly help our model explain the persistent missing inflation. If price setters interpret the recent decline in the on-the-job search rate to be a secular phenomenon, they will expect labor market conditions to remain subdued even in the very long run and hence will be even more reluctant to raise their price.[32]

[31] See Kudlyak and Faberman (2019), Faberman et al. (2019), and Abraham and Haliwanger (2019).

[32] Other stories that suggest that the fall in the on-the job search rate may be structural include Autor et al. (2017a and 2017b) who document an increase in the concentration of firms at the top of the productivity distribution. If this is the case, it may be that over time, workers have been facing less opportunities to climb the

## 4.7 The Performance of the Model in the Earlier Period

So far we have shown that the model laid out in Section 3 can overcome the failure of the baseline New Keynesian model to explain the behavior of inflation over the post-Great Recession recovery. In this section, we take on the matter of how the proposed theory of inflation fares at fitting inflation on an earlier period starting in April 1990 where our data set starts. Specifically, we look at a sample period that precedes the years in which the traditional theories of slack have clearly stopped working; that is since 2013Q1 (cf. Figure 2).

We find that the ability of our model to explain inflation dynamics is overall comparable to that implied by the traditional measures of slack in this earlier period. We do so in the simplest possible way, which is to compare how our measure of slack based on the intensity of wage competition in equation (13) performs relative to other traditional theory-based ones, using standard Phillips Curve regressions. That is, we estimate the equation

$$\pi_t = \beta \cdot slack_t + \varepsilon_t, \tag{26}$$

where $\pi_t$ is the eight-quarter moving average of the quarter-over-quarter PCE core inflation rate annualized and in deviation from two percent, which is assumed to be the long-run value for PCE core inflation.[33] The resulting inflation rate is annualized and expressed in percentage. We use the moving average as we are not interested in fitting the high-frequency swings in inflation. The variable *slack* represents different measures of labor market slack: our own, 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 Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2007), and Justiniano, Primiceri, and Tambalotti (2010). After having estimated the Phillips curve (26) for the period 1990Q2 through 2018Q4, we compute the root mean squared error (RMSE) of the different specifications over different subsamples.

All the measures of slacks deliver very similar fit of the inflation rate, attaining RMSEs bunched in an interval of around two basis points of the annualized percentage inflation rate. We conclude that our measure does comparably to other popular measures of labor market slack at fitting inflation on this sample period.

When we look at the most recent period (i.e., 2013Q1-2018Q4), the intensity of wage competition outperforms the other measures of slack by significant margins. The non-theoretical

ladder and hence ended up searching less frequently. Another empirical fact that can play a role in persistently lowering the on-the-job search rate is the decline in interstate mobility (Kaplan and Schulhofer-Wohl 2017).

[33] 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.

measure (i.e., the unemployment gap) is more than ten basis points below the explanatory performance of our measure. The gap with the other three theory-based measures of slack is even more significant, reaching 17 basis points when our measure is compared to the hours gap. This finding confirms the VAR-based results in Section 2.

# 5 Concluding Remarks

We showed that standard theories of inflation based on the New Keynesian Phillips curve fail to explain why inflation has remained subdued throughout the post-Great Recession recovery. We introduce 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 find that the model explains the recent missing inflation with the fall in the rate of on-the-job search and the associated weakening of wage competition among firms. We verify that when the on-the-job search rate is identified at micro levels using survey data, a similar fall in this rate is detected for the available years.

Our paper opens avenues for future research on the appropriate stabilization policies in the presence of interfirm competition for the employed. For instance, an important question is to understand whether monetary policy has any significant effect on the workers' willingness to search for a new job. 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., and C. J. Haltiwanger (2019): "How Tight is the Labor Market?," Federal reserve bank of chicago working paper.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2017a): "Concentrating on the Fall of the Labor Share," *American Economic Review*, 107(5), 180–185.

——— (2017b): "The Fall in the Labor Share and the Rise of Superstar Firms," Working Paper 23396, National Bureau of Economic Research.

Barlevy, G. (2002): "The Sullying Effect of Recessions," *Review of Economic Studies*, 69(1), 65–96.

Belz, S., D. Wessel, and J. Yellen (2020): "What's (Not) Up With Inflation?," The Brookings Institution.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

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.

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.

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

Elsby, M. W. L., B. Hobijn, and A. Sahin (2010): "The Labor Market in the Great Recession," *Brookings Papers on Economic Activity*, 41(1 (Spring), 1–69.

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.

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.

Jaimovich, N., I. Saporta-Eksten, H. Siu, and Y. Yedid-Levi (2020): "The Macroeconomics of Automation: Data, Theory, and Policy Analysis," mimeo.

Jaimovich, N., and H. Siu (2018): "The Trend is the Cycle: Job Polarization and Jobless Recoveries," Review of Economics and Statistics, forthcoming.

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

Kaplan, G., and S. Schulhofer-Wohl (2017): "Understanding The Long-Run Decline In Interstate Migration," *International Economic Review*, 58, 57–94.

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.

Kudlyak, M., and R. J. Faberman (2019): "The Intensity of Job Search and Search Duration," *American Economic Journal: Macroeconomics*, 11(3), 327–57.

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

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.

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.

# Appendix

In Appendix A, we show the acceptance ratio in the data. We summarize how to construct the measure of marginal costs in 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 to explain the missing inflation. We show how to work out equations (11) and (12) in the main text, which provide an analytical characterization of the surpluses in the model in Appendix F. In Appendix G, we show the robustness of our main results to change in 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 job if they receive an outside offer that makes them indifferent ($\nu$). Finally, in Appendix H, we show how we solve the model with an occasionally binding zero lower bound for the nominal interest rate.

# A Acceptance Ratio

Figure 9 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.[34] This plot shows that the acceptance ratio rapidly rose during the Great Recession. However, the acceptance ratio has steadily fallen during the recovery and is now below its pre-Great Recession average computed over the period ranging from Feb 1996-Dec 2007, which is denoted by the red dashed line.[35]

Moscarini and Postel-Vinay (2019) interpret this ratio as the acceptance ratio. Since, in their model, the fraction of accepted offers is higher when more workers are employed in low-productivity jobs, this ratio is a proxy for the degree of labor misallocation and is inversely related to inflation in their model. When this ratio 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 acceptance ratio may be due to either a high degree of misallocation or a low share of workers searching on the job. Therefore, this ratio is not always a good predictor of labor misallocation and inflation in our model. A better predictor is the empirical measure

[34] 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.

[35] The CPS data start in February 1996.

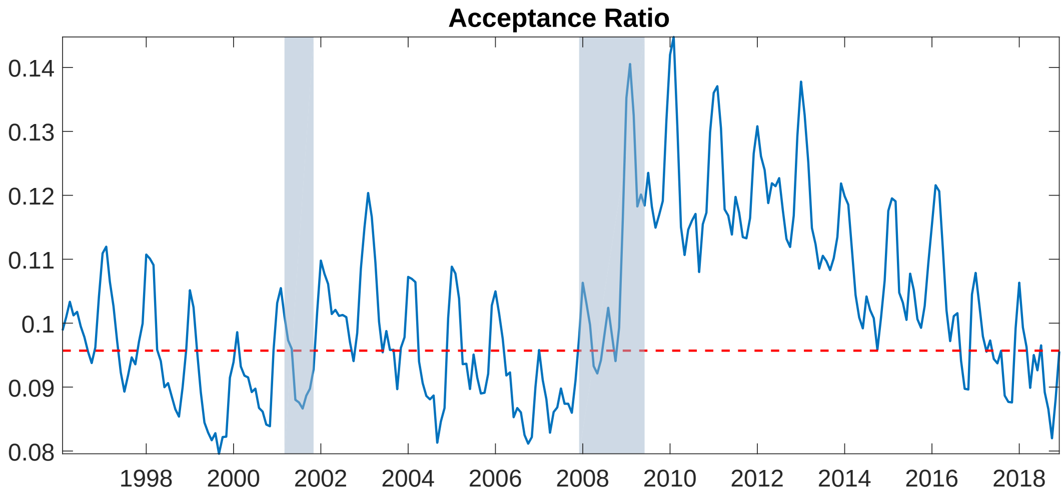


Figure 9: Acceptance Ratio. The ratio of the employment-to-employment flow rate to the unemployment-to employment flow rate. Both rates are computed by taking the three-month moving average of the CPS flow data. The red line denotes the mean of the ratio computed from 1996m2 through 2007m12. The employment-to-employment rate is corrected as proposed by Fujita, Moscarini, and Postel-Vinay (2019).

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 NK Models with S&M Frictions

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) in page 898 defines the real marginal cost as:

$$mc_t = \frac{W_t}{\alpha\left(\frac{y_t}{n_t}\right)} + \frac{c'\left(v_t\right)/q\left(\theta_t\right) - (1-\rho)\,E_t\beta_{t+1}c'\left(v_{t+1}\right)/q\left(\theta_{t+1}\right)}{\alpha\left(\frac{y_t}{n_t}\right)}, \tag{27}$$

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 (27) is the unit labor cost, i.e. the ratio of the labor cost and the marginal product of labor. The second component is 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)

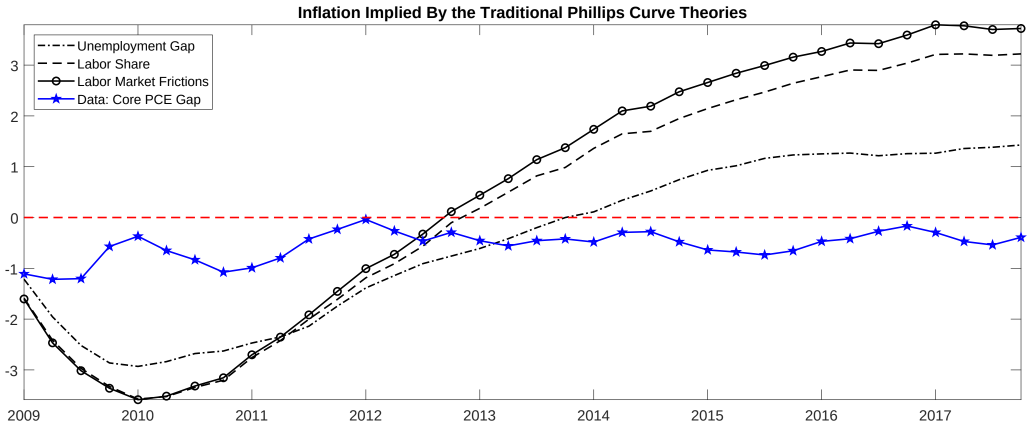


Figure 10: New Keynesian Phillips Curve with Backward-Component. Shows the inflation dynamics from 2009Q1 through 2017Q3 using three traditional theories of inflation augmented with the assumption of price indexation and the core PCE inflation gap.

show that linearizing equation (27) and rearranging, leads to the following expression:

$$\widehat{mc}_t = \hat{s}_t + \frac{1-\phi}{1-\tilde{\beta}}\left[\frac{\xi}{1-\xi}\left(\hat{h}_t - \tilde{\beta}E_t\hat{h}_{t+1}\right) + (\varepsilon_c - 1)\left(\hat{v}_t - \tilde{\beta}E_t\hat{v}_{t+1}\right) - \tilde{\beta}E_t\hat{\beta}_{t+1} - \left(1-\tilde{\beta}\right)\hat{w}_t\right], \tag{28}$$

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 Testing the Traditional New Keynesian Theories with Price Indexation

With price indexation, the New Keynesian Phillips curve becomes:

$$\pi_t = \iota\pi_{t-1} + \kappa\varphi_t + E\pi_{t+1}, \tag{29}$$

where the parameter $\iota$ controls the degree of price indexation, which affects the relative importance of the backward component of the New Keynesian Phillips curve. We can redo the same VAR-based exercise made in Section 2 in order to assess the results of that section, which were based on assuming no indexation. We set the degree of price $\iota_p$ to 0.65. The time when the inflation gap is predicted to become positive would not significantly change if one modifies this parameter within the range of values used by the empirical literature. The lagged inflation value in the first quarter of 2009 is taken from the data (i.e., it is core PCE inflation in the fourth quarter of 2018).

Figure 10 confirms the main result in the text: Price indexation just makes the drop in inflation in 2009 more pronounced and delays the period after which inflation goes above its long-run level by just three quarters. The New Keynesian Phillips curve cannot explain why we have not observed high inflation lately even if we assume price indexation.

# 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 Federal Reserve Economic Data (FRED). The labor share of income is computed as the ratio of total compensation in the non-farm business sector divided by nominal non-farm 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 (UEMP5TO14), 15 to 26 weeks (UEMP15T26), and 27 weeks and over (UEMP27OV). Primary data is 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 these time series from their trend we first remove a quadratic trend from the variables in logs,and then take the difference from their 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

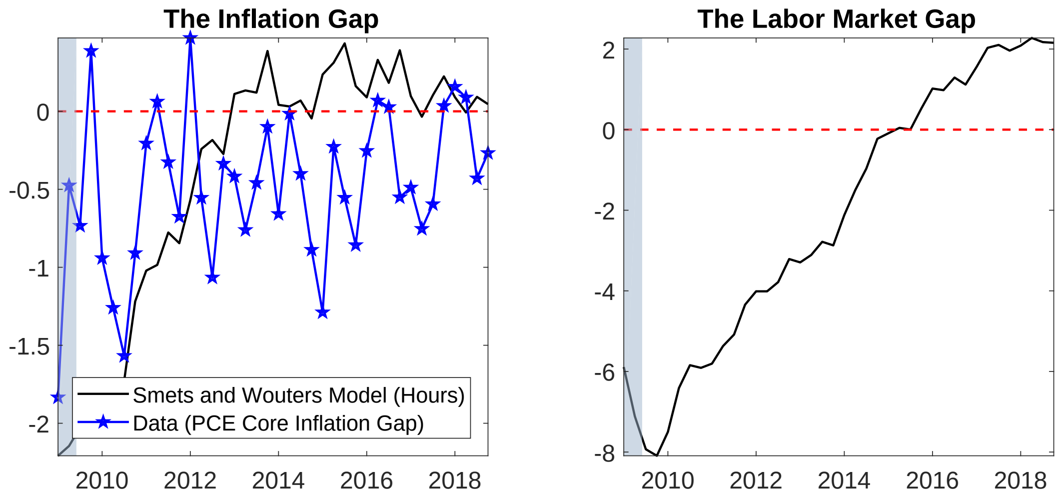


Figure 11: Left plot: the blue solid line: Inflation predicted by the model of Smets and Wouters (2007) conditional on observing the series of hours plotted on the right graph. The black starred line: the difference between the annualized quarter-to-quarter PCE core inflation rate and the ten-year-ahead PCE core inflation expectations based on the Survey of the Professional Forecasters. Right plot: Hours worked detrended using their eight-year moving average.

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 their trend as the difference from their 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 Bayesian estimation of the parameters of their model using seven observables: consumption growth, investment growth, 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 to 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 indeed the emphasis of this paper is on the apparently waning link between the labor market and inflation. The solid blue line in the right plot of Figure 11 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 of Figure 11. 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 PCE core inflation rate and the ten-year-ahead PCE core inflation expectations based on the Survey of the 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. The right plot of Figure 11, indeed, shows that the series of hours implied that the labor market becomes 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 (11), 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 reads:

$$\lambda_t V^j_{u,t} = b + \beta E_t \phi(\theta_{t+1}) \lambda_{t+1} \left[ V^j_{e,t+1} \left( w_{t+1}(j), y_{t+1}(j) \mid e^0_{t+1} = 0 \right) \right] + \beta E_t (1 - \phi(\theta_{t+1})) \lambda_{t+1} V^j_{u,t+1}, \tag{30}$$

where we let the indicator function $e^0_{t+1} = \{0, 1\}$ 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 reads:

$$\begin{aligned} \lambda_t V^j_{e,t}(w_t(j), y_t(j)) &= \lambda_t \frac{w_t(j)}{P_t} + \beta E_t \lambda_{t+1} \left\{ \delta [1 - \phi(\theta_{t+1})] V^j_{u,t+1} \right. \\ &+ \delta \phi(\theta_{t+1}) V^j_{e,t+1} \left( w_{t+1}(j), y_{t+1}(j) \mid e^0_{t+1} = 0 \right) \\ &\left. + (1-\delta) V^j_{e,t+1} \left( w_{t+1}(j), y_{t+1}(j) \mid w_t(j), y_t(j), e^0_{t+1} = 1 \right) \right\}. \end{aligned} \tag{31}$$

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 $E_t V^j_{u,t+1}$. 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 such job, expressed in units of the numeraire good is denoted by $E_t V_{e,t+1}\left[w_{t+1}(j), y_{t+1}(j) \mid e^0_{t+1}=0\right]$. With probability $1-\delta$ instead, the worker does not separate at the end of time $t$, receiving $E_t V_{e,t+1}\left[w_{t+1}(j), y_{t+1}(j) \mid w_t(j), y_t(j), e^0_{t+1}=1\right]$ 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.

$$\lambda_t V_{e,t}\left(w_t(j), y_t(j) \mid e^0_t=0\right) = b + \beta E_t \lambda_{t+1} V^j_{u,t+1} \tag{32}$$

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

$$V^j_{u,t} = \frac{b}{\lambda_t} + \beta E_t \frac{\lambda_{t+1}}{\lambda_t} V^j_{u,t+1} = V_{u,t}. \tag{33}$$

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 $\{w^*_s(y)\}^\infty_{s=t}$ denote the state-contingent contract that delivers $V^*_{e,t}(y) \equiv V_{e,t}(w^*_t, y)$. By promising to pay the contract $\{w^*_s(y)\}^\infty_{s=t}$, 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 $\{w_s(y)\}^\infty_{s=t}$. 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 $\upsilon$. 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 (10) in the text, which we report below for convenience:

$$\begin{aligned} c^f + \frac{c}{\varpi_t} &= \frac{u_{0,t}}{u_{0,t} + s_t(1-u_{0,t})} \{\xi_b S_t(y_b) + \xi_g S_t(y_g)\} \\ &+ \frac{s_t(1-u_{0,t})}{u_{0,t} + s_t(1-u_{0,t})} \left\{\xi_g \frac{l^0_{b,t}}{1-u_{0,t}} [S_t(y_g) - S_t(y_b)]\right\}. \end{aligned} \tag{34}$$

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 $\{w^*_s(y)\}^\infty_{s=t}$, 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) in the previous subsection, receives a contract from the new firm which 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 $\{w^*_s(y)\}^\infty_{s=t}$ can therefore be written as:

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

where $\varphi_t y$ is the marginal revenue product of selling $y$ units of the service to the price setters. Subtracting (33) from the above equation yields:

$$V^*_{e,t}(y) - V_{u,t} = \varphi_t y - \frac{b}{\lambda_t} + (1-\delta)\beta E_t \frac{\lambda_{t+1}}{\lambda_t} \left[V^*_{e,t+1}(y) - V_{u,t+1}\right]. \tag{36}$$

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:

$$S_t(y) = E_t\left[\sum_{\tau=0}^{\infty}(1-\delta)^{\tau}\left(\frac{\lambda_{t+\tau}}{\lambda_t}\varphi_{t+\tau}y - \frac{b}{\lambda_t}\right)\right]. \tag{37}$$

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

$$S_t(y) = y\mathcal{W}_t - \frac{b\lambda_t^{-1}}{1-\beta(1-\delta)}, \tag{38}$$

where

$$\mathcal{W}_t = \varphi_t + \beta(1-\delta)E_t\frac{\lambda_{t+1}}{\lambda_t}\mathcal{W}_{t+1}. \tag{39}$$

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, who 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 (38) we can now substitute for the surplus functions and rearrange to rewrite the free entry condition (10) as:

$$\begin{aligned} c^f + \frac{c}{\varpi_t} &= \frac{u_{0,t}}{u_{0,t}+s_t(1-u_{0,t})}\left[\mathcal{W}_t(\xi_b y_b + \xi_g y_g) - \frac{b\lambda_t^{-1}}{1-\beta(1-\delta)}\right] \\ &\quad + \frac{s_t}{u_{0,t}+s_t(1-u_{0,t})}\xi_g l_{b,t}^0\mathcal{W}_t(y_g - y_b). \end{aligned} \tag{40}$$

# G Robustness

The shaded area in the graphs of Figure 12 shows how the model's prediction of inflation changes as we varies the probability of meeting a worker that is a bad match for the firm $\xi_b$ (left) or the probability that workers switch job 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 0.25 through 0.75 and values of the parameter $\rho_s$ ranging from 0 through 0.97, which is 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 solid line and the starred red lines denote the model's predicted inflation rate and the PCE core inflation gap for the baseline calibration reported in Table 1, respectively. These lines are the same as the ones plotted in the left graph of Figure 6.

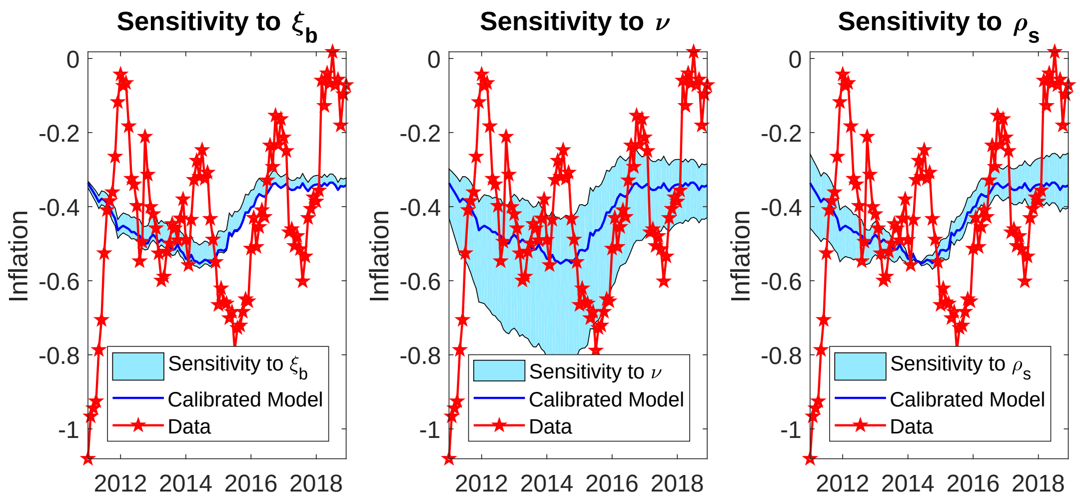


Figure 12: Robustness. Left graph: The shaded area show the sensitivity of the model's predicted year-over-year inflation rate to changes in the probability that the meeting between the worker and the firm generates a bad match ($\xi_b$). The solid blue line denotes the model's predicted year-over-year inflation rate for our baseline calibration shown in Table 1. The red starred line denotes the year-over-year inflation rate in the data (PCE core inflation) in deviations from the SPF PCE inflation expectations over the next 10 years. The middle and right graphs 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 ($\rho_s$).

# H Solving the Model with the ZLB Constraint

After being solved, our linearized model with the occasionally binding ZLB constraint in equation (18) can be represented in state-space form as follows:

$$s_t = \Gamma_0 s_{t-1} + \Gamma_1 \varepsilon_t^1 + \Gamma_2 \varepsilon_t^2 \tag{41}$$

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 (18) can be written as

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_r} \left[\left(\frac{\Pi_t}{\Pi^*}\right)^{\phi_\pi} \left(\frac{Q_t}{Q^*}\right)^{\phi_y}\right]^{1-\rho_r} + \sum_{j=0}^{k} \eta_{t-j}^j, \tag{42}$$

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 (41). These dummy shocks serve the sole purpose of enforcing the ZLB constraint (i.e., prevent agents to expect 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 notes that the matrix $\Gamma_1$ is a matrix with $k+1$ columns.

As explained in the text, the shocks are obtained by basically 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. Assuming that the matrix $Z\Gamma_2$ is invertible (as it is in our case), these inversion allows us to retrieve the sequence of shocks $\varepsilon_t^2$ that identically explain 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:

$$\Delta\varepsilon_t^1 = \left(\Gamma_1^{(0:k)}\right)^{-1}\left[-\ln R_* - \Gamma_0^{(0:k)} s_{t-1} - \Gamma_1^{(0:k)}(i)\cdot\Psi(i)\,\varepsilon_t^1(i) - \Gamma_2^{(0:k)}\varepsilon_t^2\right] \tag{43}$$

   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),

$$\Gamma_0^{(0:k)} s_{t-1} + \Gamma_1^{(0:k)}\cdot\Psi(i)\,\varepsilon_t^1(i+1) + \Gamma_2^{(0:k)}\varepsilon_t^2 > -\ln R_* \tag{44}$$

   We adjust the diagonal matrix of zeros and ones, $\Psi(i+1)$, so that the set of horizons at which the ZLB is binding are 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 and $i = i+1$ and go to Step 2, 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:

$$s_t = \Gamma_0 s_{t-1} + \Gamma_1\varepsilon_t^1(i+1) + \Gamma_2\varepsilon_t^2. \tag{45}$$

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

The $s_t$ coming from equation (45) is the vector containing the model predicted value of the state variables, which is used to generate all the empirical results of the paper.