Slow Recoveries and Unemployment Traps: Monetary Policy in a Time of Hysteresis*

Sushant Acharya Julien Bengui Keshav Dogra Shu Lin Wee

March 8, 2021

Abstract

We analyze monetary policy in a model where temporary shocks can permanently scar the economys productive capacity. Workers lose skill while unemployed and are costly to retrain, generating multiple steady-state unemployment rates. Following a large shock, unless monetary policy acts aggressively and quickly enough to prevent a significant rise in unemployment, hiring falls to a point where the economy recovers slowly at best – at worst, it falls into a permanent unemployment trap. Monetary policy can only avoid these outcomes if it commits in a timely manner to more accommodative policy in the future. Timely commitment is essential as the effectiveness of monetary policy is state dependent: once the recession has left substantial scars, monetary policy cannot speed up a slow recovery, or escape from an unemployment trap.

JEL Classification: E24, E3, E5, J23, J64

Keywords: hysteresis, path dependence, monetary policy, multiple steady states, skill depreciation

*Acharya: Bank of Canada & CEPR, sacharya@bankofcanada.ca; Bengui: Bank of Canada & CEPR, jbengui@bankofcanada.ca; Dogra: FRB NY, keshav.dogra@ny.frb.org; Wee: Bank of Canada, shulinwee@bankofcanada.ca.
The authors are grateful to Luca Fornaro, Luigi Iovino, Ben Johannsen and Edouard Schaal for insightful comments and discussions. We also thank Jess Benhabib, Javier Bianchi, Marco Del Negro, Roger Farmer, Jordi Gali, Marc Giannoni, Todd Keister, Rasmus Lentz, David Ratner and seminar participants at Boston College, Duke University, Rutgers University, University of Pittsburgh, University of Wisconsin - Madison, Federal Reserve Board of Governors, Federal Reserve Bank of New York, SED 2018, the 7th BI-CEPR conference on Low Inflation and Wage dynamics: Implications for Monetary policy and Financial stability, the Macroeconomic System Committee meeting and the 2019 CEPR ESSIM for useful discussion and comments. The paper was previously circulated under the title "Escaping Unemployment Traps". The views expressed in this paper are those of the authors and do not necessarily represent those of the Bank of Canada, Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

In the aftermath of the global financial crisis, economic activity remained subdued, suggesting that the world economy may have settled on a lower growth trajectory than the one prevailing before 2007. Some observers have attributed this sluggish growth to permanent, exogenous structural changes - either permanently lower productivity growth (Gordon, 2015) or *secular stagnation*.[1] An alternative explanation is that large, temporary downturns can themselves permanently damage an economy's productive capacity. This hysteresis view, according to which changes in current aggregate demand can have a significant effect on future aggregate supply, dates back to the 1980s but recently underwent a surge of interest in the wake of the Great Recession (e.g., Yellen, 2016). While the two sets of explanations may be observationally similar, they have very different normative implications. If exogenous structural factors drive slow growth, countercyclical policy may be unable to resist or reverse this trend. In contrast, if temporary downturns themselves lead to persistently or permanently slower growth, then countercyclical policy, by limiting the severity of downturns, may have a role to play to avert such adverse developments.[2]

In this paper, we present a theory in which hysteresis might occur and countercyclical monetary policy can moderate its impact if timed appropriately. In our model, hysteresis can arise because workers lose human capital whilst unemployed and unskilled workers are costly to retrain, as in Pissarides (1992). In the presence of nominal rigidities and a zero lower bound (ZLB) constraint on monetary policy, large adverse fundamental shocks can cause recessions whose legacy is persistent or permanent unemployment. Under this setting, the timing of monetary policy matters significantly for long-term outcomes. Timely commitment to future accommodative policy early in a recession can prevent hysteresis from taking root and enable a swift recovery. In contrast, delayed monetary policy interventions may be powerless to bring the economy back to full employment.

Our environment is an economy with downwardly sticky nominal wages and labor search frictions. Human capital depreciates during unemployment spells and unskilled workers are costly to retrain. These features generate multiple steady states. One steady state is a high pressure economy: job finding rates are high, unemployment is low and job-seekers are highly skilled. While tight labor markets - by improving workers' outside options - cause wages to be high, firms still find job creation attractive, as higher wages are offset by low average training costs when job-seekers are mostly highly skilled. The economy, however, can also be trapped in a low pressure steady state. In this steady state, job finding rates are low, unemployment is high, and many job-seekers are unskilled as long unemployment spells have eroded their human capital. Slack labor markets lower the outside options of workers and drive wages down, but hiring is still limited as firms find it costly to retrain these workers. Crucially, this is an *unemployment trap* - an economy near the low pressure steady state can

[1] Here by *secular stagnation* we refer to the literature arguing that a chronic excess of global savings relative to investment has depressed equilibrium real interest rates. This imbalance has been attributed to permanent changes in either borrowing constraints, supply of safe assets, demographics, inequality or monopoly power. See, for example, Eggertsson and Mehrotra (2014) and Caballero and Farhi (2017), among many others.

[2] In the words of Yellen (2016): *"...hysteresis would seem to make it even more important for policymakers to act quickly and aggressively in response to a recession, because doing so would help to reduce the depth and persistence of the downturn, thereby limiting the supply-side damage that might otherwise ensue."*

never self-correct and return to a high pressure state.

Nominal wage rigidities and a lower bound constraint on monetary policy can enable temporary shocks to permanently move the economy from a high pressure steady state into an unemployment trap. Consider an environment where the monetary authority pursues what we term *conventional* monetary policy (CMP): it aims to implement the highest employment level consistent alongside zero nominal wage inflation. A large but temporary increase in households' discount factor raises desired savings, pushing the real interest rate below zero. Monetary policy tries to accommodate the excess demand for savings by lowering the nominal interest rate below zero, but is constrained by the ZLB. Consequently, current prices are forced to adjust downwards, as households' demand for current consumption relative to the future declines. Under downward nominal wage rigidity, the decline in prices causes real wages to rise, and hiring to fall, lengthening average unemployment duration and increasing the incidence of skill loss. Deterioration in average skill quality among the unemployed in turn raises the effective cost of job creation, discouraging vacancy posting and slowing the economy's return to the high pressure steady state even after the shock has abated. In the event of a large enough shock, the economy may be pushed into an unemployment trap from which it is powerless to escape.

The transition to an unemployment trap following a large adverse shock can be avoided if, rather than pursuing CMP, the monetary authority instead engages in what we term *unconventional* monetary policy (UMP). By instead committing to temporarily higher inflation after the liquidity trap has ended, the monetary authority can mitigate both the initial rise in unemployment, and its persistent (or permanent) negative consequences. UMP, however, is only effective if it is implemented early in the downturn, before the recession has left substantial scars. Once the skill composition of the unemployed has significantly worsened following the shock, monetary policy cannot undo the high cost of hiring through the promise of higher future prices. With nominal wages free to adjust upwards, any attempt to generate price inflation is met by nominal wage inflation, leaving real wages unaffected. Thus, once hysteresis has taken root, monetary policy cannot undo it. In such cases, fiscal policy, in the form of hiring or training subsidies, is necessary to engineer a recovery.

One might think that it is more natural to address hysteresis through such fiscal policies more generally. In our model, as in the standard New Keynesian (NK) model, an appropriately rich set of fiscal instruments would wholly obviate the need for countercyclical monetary policy (Correia et al., 2008). In reality, fiscal policy is imperfect and slow to respond to a downturn, leaving monetary policy as the first responder when it comes to countercyclical stabilization. The NK literature on stabilization policy concentrates on monetary rather than fiscal policy for precisely this reason. Given the limitations of fiscal policy, it is important to know if and when monetary policy can prevent hysteresis or mitigate its effects. Moreover, even if fiscal policy in the form of hiring subsidies were used to engineer a recovery, such a policy would not be costless as households must forgo consumption when more resources are allocated towards subsidizing job creation. Unconventional monetary policy – which prevents any initial rise in unemployment – avoids these costs altogether, suggesting that fiscal policy may not even be a good substitute for timely monetary policy action.

Our paper underscores the importance of *timely* monetary policy accommodation. In the presence of hysteresis, a failure to deliver stimulus early on in a recession can have irreversible costs. This contrasts with standard NK models, in which accommodative policies are equally effective at any point in a liquidity trap (Eggertsson and Woodford, 2003). In fact, these models predict that while overly tight policy may be costly in the short-run, it has no long-run consequences, since temporary shocks have no permanent effects in stationary models. Consequently, timeliness of monetary accommodation is not particularly relevant in these models, as delaying accommodation does not reduce welfare in the long run. Our model instead focuses on a monetary economy with multiple steady states. Thus, monetary policy can affect not just fluctuations around a steady state, but also the level of steady state activity.[3]

Our focus on multiple steady states also distinguishes our paper from recent work on the persistent effects of recessions (Benigno and Fornaro, 2017; Schmitt-Grohe and Uribe, 2017). These papers study economies which can switch to a bad equilibrium featuring permanently low or negative inflation, a binding ZLB, and high unemployment. This bad equilibrium is the result of self-fulfilling pessimistic beliefs; equally, self-fulfilling optimism can return the economy to the good equilibrium. Our analysis differs in two ways. First, in our model, high unemployment can persist even after monetary policy is no longer constrained by the ZLB. Second, it features path dependence: optimistic beliefs cannot return the economy back to full employment if the economy is stuck in an unemployment trap, nor can they speed up a slow recovery. This is because dynamics in our economy are driven by a slow-moving state variable - the fraction of unskilled job-seekers. Even if a swift recovery is anticipated, this does not induce firms to hire and train relatively unskilled job-seekers today. In fact, firms postpone hiring, preferring to wait until there are more skilled job-seekers. Since hiring falls, the skill composition of job-seekers actually worsens and the firms' optimism is self-defeating. Since self-fulfilling optimism cannot escape the trap ex post, it is all the more important to avoid it ex ante.

The remainder of the paper is structured as follows. Next we discuss related literature. Section 2 presents the model economy. Section 3 characterizes steady states and equilibrium under flexible wages. Section 4 introduces nominal rigidities, and studies how demand shocks can cause slow recoveries or permanent stagnation. Section 5 presents some extensions and discussions. Section 6 concludes.

Related literature Our paper relates most closely to a small number of recent studies analyzing hysteresis and monetary policy in the presence of nominal frictions. Benigno and Fornaro (2017) present a model in which pessimism can drive the economy to the ZLB, reducing firms' incentive to innovate and giving rise to persistent or permanent slowdowns. A commitment to alternative monetary policy rules (or subsidies to innovation) can help avoid or exit such *stagnation traps*. While we study a different channel through which hysteresis might operate, and focus on unemployment rather than output hysteresis, our results resonate with theirs. A key difference is that in our model, a commitment to accommodative monetary policy can only avoid an unemployment trap if it is implemented swiftly; if the economy is already stuck in such a trap, monetary policy is of little help. Bianchi et al. (2014) also find that declines in R&D during recessions can explain persistent effects of cyclical shocks on growth, while Garga and Singh (2016) study the conduct of optimal monetary policy in a model

[3]This does not mean that monetary policy can manipulate a long-run trade-off between inflation and unemployment. In our baseline model, once the economy has converged to a particular steady state unemployment rate, monetary policy is powerless to reduce unemployment below this rate.

embedding this feature. Laureys (2014) studies monetary policy in an environment, similar to ours, where skills depreciate during unemployment spells, but focuses on linear dynamics around a unique steady state. In the same vein, Galí (2016) studies optimal monetary policy in a NK model where insider-outsider labor markets can generate hysteresis. In these two papers, however, temporary shocks and policy mistakes have persistent but not irreversible effects. By studying an environment where temporary shocks can cause irreversible damage, we are able to stress the distinctive role of timeliness in monetary policy action.

Our analysis also contributes to the broader theoretical literature studying hysteresis, which has largely abstracted from nominal rigidities. Drazen (1985) argues that the loss of human capital due to job loss in recessions can lead to delayed recoveries. Schaal and Taschereau-Dumouchel (2016) show that a labor search model with aggregate demand externalities can generate additional persistence in labor market variables. Similarly, in Schaal and Taschereau-Dumouchel (2015), large recessions frustrate coordination on a high-activity equilibrium, allowing temporary shocks to cause quasi-permanent recessions. Our model instead draws on Pissarides (1992), who argued that skill depreciation can give rise to multiple steady states. Sterk (2016) studies a quantitative version of Pissarides (1992)'s model and argues that it can account for the behavior of job finding rates in the United States. Relative to our work, all these studies consider purely real models. As such, they do not address the question we are interested in - namely, how monetary policy should be conducted in the presence of hysteresis.

On the empirical side, a large literature finds evidence of drops in productive capacity after recessions. Dickens (1982) finds that recessions can permanently lower productivity; Haltmaier (2012) finds that trend output falls by 3 percentage points on average in developed economies four years after a pre-recession peak. Using cross-country data, Martin et al. (2014) find that severe recessions have a sustained and sizable negative impact on output. Similarly, Ball (2014) finds that countries with a larger fall in output during the Great Recession experienced a larger decline in *potential output*. Within the U.S., Yagan (2017) finds that states exposed to larger unemployment shocks in 2007 experienced significantly lower employment rates in 2015. Song and von Wachter (2014) find that the persistent decline in employment following job displacement is larger during recessions, suggesting that a spike in job destruction rates can persistently affect unemployment.

Our work also contributes to the large literature exploring how downward nominal wage rigidity can exacerbate unemployment outcomes during a recession. While the literature has largely examined how downward nominal wage rigidities can raise the spectre of layoffs during recessions (e.g. Murray 2019), our paper highlights how such rigidities can discourage job creation by raising the effective cost of hiring. A large and growing empirical literature supports our assumption that nominal wages are downwardly rigid. Using payroll data, Grigsby et al. (2019) find that only 2.4% of all workers observe a nominal wage cut during a year and that wages of new hires do not appear to be more flexible than those of incumbent workers. Further, at the height of the Great Recession, they find that only 6% of workers observed a nominal wage cut. Barattieri et al. (2014) and Fallick et al. (2020) find similar evidence that nominal wage cuts are infrequent using data from the Survey of Income and Program Participation (SIPP) and the Bureau of Labor Statistics (BLS), respectively. Finally, using data on the wages of new hires at the *job level*, Hazell and Taska (2019) find that nominal wage cuts account for only 9% of the adjustments in new hires' posted wages.

Beyond the literature on hysteresis and downward nominal wage rigidities, our analysis also connects to a few recent developments in monetary economics. Like us, Dupraz et al. (2017) study a *plucking model* in which downward nominal wage rigidity gives rise to asymmetric effects of monetary policy: while deflation can lead to an increase in real wages and a fall in hiring, inflation has limited effects on unemployment. In their model, this asymmetry increases the costs of business cycles despite shocks having at most a temporary effect. Our analysis suggests that the costs associated with this asymmetry become amplified when combined with hysteresis: temporary deflation can lead to permanently higher unemployment and deterioration in the skill composition of the unemployed, both of which cannot be reversed by higher inflation in the future. This underscores the relevance of timely monetary policy to stabilize employment, even at the cost of compromising price stability. Our result resonates with Acharya et al. (2020) and Berger et al. (2016), who find that monetary policy should prioritize output and employment stabilization over price stability when households are imperfectly insured. Our analysis provides another reason why employment fluctuations might have higher costs, and warrant more attention.

Finally, our paper relates to the secular stagnation literature. Eggertsson and Mehrotra (2014) and Caballero and Farhi (2017) present models in which the market clearing interest rate is persistently or permanently negative, leading to persistently low output, as the ZLB prevents nominal rates from falling to clear markets. In such situations, a permanent change in fiscal or monetary policy is typically required to prevent stagnation. We share this literature's concern with long run outcomes, but consider a different mechanism: in our model *temporary* falls in market clearing interest rates have permanent effects, which *temporary* monetary accommodation can prevent.

2 The Model Economy

We start by presenting a benchmark economy with labor market frictions à la Diamond-Mortensen-Pissarides (DMP) and no nominal rigidities. Time is discrete and there is no uncertainty. The only addition to the standard DMP model is that we assume that workers can lose skill following an unemployment spell.

Workers There is a unit mass of risk-neutral ex ante identical workers with discount factor β. Workers trade nominal bonds which pay a nominal return of $1 + i_t$. Workers can either be employed or unemployed. We denote the mass of employed workers as n and unemployed as $u = 1 - n$. Unemployed workers produce $b > 0$ as home production. The stock of employed workers evolves as:

$$n_t = [1 - \delta(1 - q_t)]\, n_{t-1} + q_t u_{t-1}, \tag{1}$$

where δ is the exogenous rate at which workers are separated from their current jobs and q_t is the job finding rate. (1) implies that a worker separated at the beginning of period t can find another job within the same period. Next, let $\mathbb{W}_t$ denote the value of an employed worker and $\mathbb{U}_t$ denote the value of an unemployed worker at time t. These can be expressed as follows:

$$\begin{aligned}
\mathbb{W}_t &= \omega_t + \beta\Big\{\left[1-\delta(1-q_{t+1})\right]\mathbb{W}_{t+1} + \delta(1-q_{t+1})\mathbb{U}_{t+1}\Big\}, && (2)\\
\mathbb{U}_t &= b + \beta\Big\{q_{t+1}\mathbb{W}_{t+1} + (1-q_{t+1})\mathbb{U}_{t+1}\Big\}, && (3)
\end{aligned}$$

where ω_t denotes the real wage at date t.

Labor market In the spirit of Pissarides (1992), we assume that a worker who gets separated from her job and who fails to transition back to employment by the end of a period immediately loses the skills that she acquired while employed. That is, any worker unemployed for at least 1 period becomes *unskilled*. Because unskilled workers produce zero output when matched with a firm, a firm that hires an unskilled worker must pay a training cost $\chi > 0$ to use that worker in production. Once the firm trains the worker, she remains *skilled* until the next unemployment spell of at least 1 period. Let μ_t denote the fraction of unskilled workers in the pool of job-seekers (l_t) at date t. This fraction is defined as:

$$\mu_t = \frac{u_{t-1}}{l_t} \equiv \frac{u_{t-1}}{1-(1-\delta)(1-u_{t-1})}. \tag{4}$$

(4) shows that a higher level of unemployment in the past corresponds to a higher fraction of unskilled job-seekers. As such, there is a one-to-one mapping between u_{t-1} and μ_t.

Matching technology Search is random. The number of successful matches m_t between job-seekers l_t and vacancies v_t is given by a CRS matching technology $m(v_t, l_t)$. We define market tightness θ_t as the ratio of vacancies to job-seekers. The job finding probability of a job-seeker, q_t, and the job filling probability of a vacancy, f_t, are then given by:

$$q(\theta_t) = \frac{m(v_t, l_t)}{l_t} \qquad \text{and} \qquad f(\theta_t) = \frac{m(v_t, l_t)}{v_t} = \frac{q(\theta_t)}{\theta_t}. \tag{5}$$

Firms A representative CRS firm uses labor as an input to produce the final good. The production function is given by $y_t = An_t$ where $A > b$ is aggregate productivity and n_t is the number of employed workers in period t. A firm must incur a vacancy posting cost of $\kappa > 0$ and an additional training cost of χ for each unskilled worker hired. A firm with n_{t-1} workers at the beginning of period t chooses vacancies (taking wages as given) to maximize lifetime discounted profit:

$$\mathbb{J}_t = \max_{v_t \geq 0}(A-\omega_t)n_t - (\kappa + \chi\mu_t f_t)v_t + \beta\mathbb{J}_{t+1} \qquad \text{s.t.} \qquad n_t = (1-\delta)n_{t-1} + f_t v_t,$$

where ω_t is the wage paid to all workers. Importantly, the total cost of job creation depends on the skill composition of job-seekers. Since the firm pays a cost χ to train each unskilled job-seeker it hires, the effective average cost of creating a job is increasing in the fraction of unskilled job-seekers μ_t. From (4), μ_t depends on past unemployment rates, making the cost of job creation *increasing* in the unemployment rate. The value of a filled vacancy, $J_t = \partial \mathbb{J}_t / \partial n_t$, can be written as

$$J_t = A - \omega_t + \beta(1-\delta)J_{t+1}. \tag{6}$$

Free entry of vacancies implies

$$J_t \leq \frac{\kappa}{f_t} + \chi\mu_t \ , \ \theta_t \geq 0 \ , \ \textit{with at least one strict equality.} \tag{7}$$

Resource constraint The resource constraint can be written as

$$c_t = An_t + b(1-n_t) - (\kappa + \chi\mu_t f_t)\, v_t.$$

Wage and price determination While we ultimately seek to analyze the conduct of monetary policy in an environment with nominal wage rigidities, it is useful to first study a flexible wage benchmark, in which wages are simply determined by Nash bargaining every period. Because bargaining occurs after all hiring and training costs have been paid, all workers are paid the same wage.[4] Formally, the Nash bargaining problem is $\max_{\omega_t} J_t^{1-\eta} (\mathbb{W}_t - \mathbb{U}_t)^\eta$ where $\eta \in [0,1)$ denotes the bargaining power of the workers. The Nash-bargained wage is

$$\omega_t^* \quad = \quad \eta A + (1-\eta)b + \beta(1-\delta)\eta q_{t+1} J_{t+1}. \tag{8}$$

Crucially, an increase in next period's job finding rate puts upward pressure on the wage because it increases the worker's outside option. Substituting (8) into (6) yields

$$J_t = a + \beta(1-\delta)(1-\eta q_{t+1})J_{t+1}, \tag{9}$$

where $a \equiv (1-\eta)(A-b)$. (9) implies that the value of a filled vacancy to a firm lies in the interval: $J_t \in [J_{min}, J_{max}]$, for $J_{min} \equiv a/[1-\beta(1-\delta)(1-\eta)]$ and $J_{max} \equiv a/\left[1-\beta(1-\delta)\right]$.[5] It also implies that an increase in the job finding rate at every future date, through an upward pressure on the wage, results in a smaller profit to the firm and thus a lower J_t. In this flexible wage benchmark, the classical dichotomy holds and the price level does not affect real allocations. Thus, it is not necessary to describe the conduct of monetary policy. Equilibrium dynamics in the benchmark economy is completely characterized by (7), (9) and the law of motion for μ_t, which is given by:

$$\mu_{t+1} \quad = \quad \frac{1-q(\theta_t)}{1+(1-\delta)[1-q(\theta_t)-\mu_t]}. \tag{10}$$

For analytical tractability, in Sections 3 and 4, we assume a particular form for the matching function, $m_t = \min\{v_t, l_t\}$, which implies $q(\theta_t) = \min\{\theta_t, 1\}$, $f(\theta_t) = \min\{1, 1/\theta_t\}$. In particular, it implies that the short side of the market matches with probability 1. We refer to the case with $\theta_t < 1$ as the *slack*

[4]Both skilled and unskilled workers have the same outside option since training costs are sunk at the time of bargaining and all job-seekers have the same probability of finding a job.

[5]J_{min} is the lowest value of a filled vacancy and is achieved when firms expect $q_t = 1$ forever. Conversely, J_{max} is the value of a filled vacancy when firms expect $q_t = 0$ forever (labor markets are expected to be the slackest forever).

labor market regime and the one with $\theta_t \geq 1$ as the *tight labor market regime*. In Appendix L, we explore the quantitative implications of our model when we use a more standard matching function, such as the CES matching function.

3 Flexible wage benchmark

Our goal is to study how temporary shocks can scar the economy permanently depending on the conduct of monetary policy. Permanent scarring is possible in our model because it features multiple steady states, and shocks can push the economy from one steady state to another. In this section, we explain why multiple steady states exist in our economy and characterize dynamics in the flexible wage benchmark.

3.1 Steady states

In our model, multiplicity of steady state unemployment rates arises naturally because workers lose skill while unemployed and firms must pay a cost to train unskilled workers. Consider an economy with high unemployment. Since average unemployment duration is high, the fraction of unskilled job-seekers is high. Consequently, firms must spend more on training workers, which discourages them from creating vacancies, even though slack labor markets lower workers' outside options and drive down wages. Thus, a high unemployment rate is self-sustaining. Conversely, when unemployment is low, mean unemployment duration is low and the fraction of skilled job-seekers is high. While wages are high since tight labor markets improve workers' outside options, firms still post vacancies because expected training costs are low as most job-seekers are skilled. This in turn sustains low unemployment. Given our Leontief matching function, the low unemployment steady state corresponds to zero unemployment. Such a steady state exists under the following assumption.

Assumption 1. *Vacancy posting costs are low enough:* $\kappa < J_{min}$.

The law of motion for employment (1) implies that full employment ($n = 1$) requires $q = 1$ (and $f = 1/\theta \leq 1$); job-seekers are on the short side of the market, and always find a job within one period. Skill depreciation never occurs, and the law of motion for the skill composition (10) implies $\mu = 0$. Thus, the effective cost of hiring a worker is simply κ/f and the job creation condition (7) becomes $J_{min} = \kappa\theta^{\text{fe}}$ in steady state, where θ^{fe} denotes the labor market tightness associated with full employment. Assumption 1 ensures that this equation has a solution featuring $\theta^{\text{fe}} > 1$.

While skill depreciation *can* generate multiple steady states, whether it in fact does so depends on the strength of the scarring effects of unemployment (measured by χ) and the sensitivity of wages to workers' outside options (measured by η). The following assumption ensures that both forces are strong enough such that in addition to the full employment steady state, there exist additional interior steady states featuring higher unemployment.

Assumption 2. *The training cost* χ *is neither too small nor too large, i.e.,* $\chi \in (\underline{\chi}, J_{max} - \kappa)$, *and the workers' bargaining power is not too small,* $\eta > \underline{\eta}$. *The thresholds* $\underline{\eta}$ *and* $\underline{\chi}$ *are defined in Appendix B.*

Appendix B shows that $\kappa + \chi < J_{max}$ ensures that training costs are not too large, so that the worst steady state features a positive level of employment.[6] The remaining elements of Assumption 2 ensure that two interior steady states with unemployment exist (in addition to the full employment steady state). From the law of motion for employment (1), at any interior steady state ($n < 1$), firms are on the short side of the labor market ($q < 1$). This implies that there is some skill depreciation ($\mu > 0$), since from the law of motion for the skill composition (10), we have $q = 1 - \mu < 1$ in steady state. At an interior steady state, the job creation condition (7) becomes[7]

$$\frac{a}{1 - \beta(1-\delta)\left[1 - \eta(1-\mu)\right]} = \kappa + \chi\mu. \tag{11}$$

The left-hand side (LHS) of (11) is the value of a filled vacancy with $q = 1 - \mu$, while its right-hand side (RHS) is the cost of creating a job. (11) describes a quadratic in μ, which has at most two solutions. Appendix B shows that Assumption 2 guarantees that economically meaningful solutions to this equation exist. A high bargaining power η increases the sensitivity of wages and profits to labor market conditions. When unemployment is low, wages are high because the worker's outside option is relatively favorable. Firms are willing to tolerate high wages because training costs are low. When labor markets are slack and unemployment is high, workers are relatively unskilled and expensive to train; firms are willing to pay the high training costs because wages are relatively low.

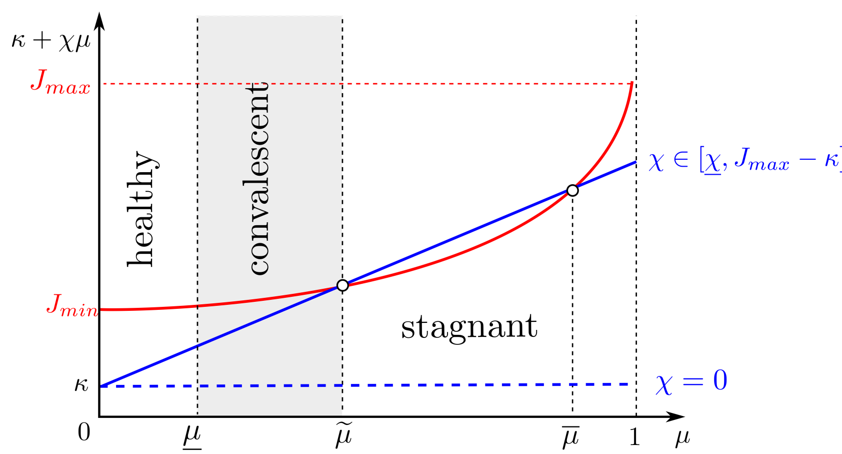


Figure 1. Multiple steady states: The red curve depicts the LHS of (11), the blue line depicts the RHS.

Figure 1 graphically depicts the arguments above. The red curve plots the LHS of (11), while the blue line plots its RHS for different values of χ. When χ is too low, the two curves do not intersect and there are no interior steady states. If χ is too high, the blue line lies above the red curve at $\mu = 1$ and there exists a zero-employment steady state, in violation of Assumption 2. When χ is in the

[6] If there was zero employment in steady state, everyone would be unskilled ($\mu = 1$) and labor markets would be completely slack ($q = 0$). Since wages are the lowest they can be, the value of a filled vacancy is J_{max}, the highest it can be. Since all workers are unskilled, the effective cost of hiring a worker is now $\kappa + \chi$. $\kappa + \chi < J_{max}$ ensures that at even such a high level of μ, a firm would find it profitable to post some vacancies, ruling out the uninteresting possibility of a zero employment steady state. Qualitatively, none of our results would change if we allowed for a zero employment steady state.

[7] In this section and in what follows, it will be convenient to work with the fraction of unskilled workers μ rather than the unemployment rate u as the state variable of interest. Equation (4) defines a one-to-one map between μ_t and u_{t-1}.

appropriate range, then there are two interior steady states, $\widetilde{\mu}$ and $\overline{\mu}$ (with $\widetilde{\mu} < \overline{\mu}$). Finally, recall that there is always a full employment steady state at $\mu = 0$. The three steady states are associated with different degrees of market tightness, and accordingly, with different levels of wages: wages at the full employment steady state are higher than at the moderate unemployment steady state $\widetilde{\mu}$, which in turn are higher than at the high unemployment steady state $\overline{\mu}$.

Multiple steady states under alternative search specifications Our model is highly stylized. In reality, firms may be able to distinguish between skilled and unskilled applicants – either when interviewing them or even before, when they specify the job-requirements when posting a vacancy. In Appendix K, we extend the model in various dimensions and show that skill depreciation generally results in the existence of multiple steady states, even when firms can test workers' skills, or post vacancies in segmented markets.

3.2 Dynamics

Next, we characterize the transitional dynamics of the economy starting from any $\mu_0 \in [0, 1]$. While we have thus far not introduced any aggregate shocks which would move the economy away from a steady state, we will do so in Section 4. For now, we can think of the experiment as studying the evolution of the economy after past shocks have moved it to a point μ_0. The subsequent evolution of the economy can be described by a mapping $\mu_{t+1} = \mathcal{M}(\mu_t)$.

As indicated in Figure 1, the state space can be partitioned into 3 regions, depending on the initial fraction of unskilled job-seekers μ: (i) a *healthy region* featuring low unemployment and a highly skilled workforce (low μ), (ii) a *convalescent region* featuring moderate levels of unemployment and a moderately skilled workforce (intermediate level of μ); and finally (iii) a *stagnant region* with high unemployment and a largely unskilled workforce (high μ). Dynamics differ between these three regions, as we now describe.

Healthy region If the economy starts in the healthy region, defined as $\mu \in [0, \underline{\mu}]$ where $\underline{\mu} \equiv (J_{min} - \kappa)/\chi < \widetilde{\mu}$, then labor markets are tight and the economy immediately converges back to the full employment steady state, as formalized in the following proposition.

Proposition 1. *Suppose the economy starts in the healthy region, $\mu_0 < \underline{\mu}$. Then the economy converges to the full employment steady state in one period, $\theta_t = (J_{min} - \chi\mu_0)/\kappa$ for $t = 0$, and $\theta_t = J_{min}/\kappa > 1$, $n_t = 1$, $\mu_t = 0$ for $t \geq 1$. Furthermore, the value of a filled vacancy and wages are always at their full employment steady state level: $J_t = J_{min}$ and $\omega_t = \omega^*_{\text{fe}} \equiv \eta A + (1 - \eta)b + \beta\eta(1 - \delta)J_{min}$ for all $t \geq 0$.*[8]

Proof. See Appendix C. □

Intuitively, when the unemployment rate is very low, average skill quality of job-seekers is very high. Hence, low training costs make it attractive for firms to post enough vacancies to absorb all

[8] The equilibrium is unique, except in the knife-edge case where $\mu = \underline{\mu}$, where there also exists other equilibria in which the economy returns to the full employment steady state in 2 periods instead of 1. Note, however, that in these equilibria the value of a filled vacancy and the real wage also satisfy $J_t = J_{min}$, $\omega_t = \omega^*_{\text{fe}}$ for all $t \geq 0$.

job-seekers, despite the high wages associated with tight labor markets in the present and future. Consequently, unemployment duration is short (everyone finds a job by the end of the first period), and the skill quality of the workforce remains high. While we have not yet introduced any shocks, one interpretation is that the full employment steady state is stable with respect to shocks which only cause small deteriorations in the average skill composition of job-seekers. In particular, if μ_0 rises to a level in the interval $(0, \underline{\mu}]$, the effect of the shock is immediately reversed as job-seekers are still largely skilled, and firms are willing to post enough vacancies to hire and retrain all job-seekers on the spot. As long as $\mu_0 < \underline{\mu}$, $\theta_0 > 1$ and the economy immediately returns to full employment: $\mu_1 = \mathcal{M}(\mu_0) = 0$.

Convalescent region If the economy starts in the convalescent region, $\mu_0 \in (\underline{\mu}, \widetilde{\mu})$, it eventually returns to full employment, but does not do so instantaneously:

Proposition 2 (Dynamics in the convalescent region)**.** *For β sufficiently close to 1, there exists a unique strictly increasing sequence $\{\mu^n\}_{n=0}^{\infty}$ with $\mu^0 \equiv \underline{\mu}$ and $\lim_{n\to\infty} \mu^n = \widetilde{\mu}$, such that if $\mu_0 \in I^n \equiv (\mu^{n-1}, \mu^n]$, the economy reaches the healthy region in n periods and reaches the full-employment steady-state in $n+2$ periods, i.e., $\mu_n = \underline{\mu}$, $\mu_{n+1} \in (0, \underline{\mu})$ and $\mu_{n+2} = 0$. Furthermore:*

1. *Recoveries can be arbitrarily long: As $\mu_0 \to \widetilde{\mu}$, the time it takes for the economy to return to the healthy region tends to infinity.*[9]

2. *Recoveries can be arbitrarily slow: If μ_0 is close to $\widetilde{\mu}$, then μ declines very slowly early on in the recovery.*[10]

Proof. See Appendix D. □

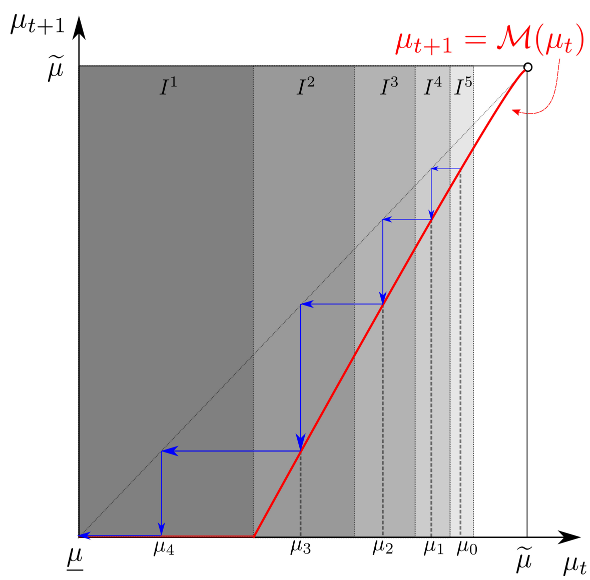


Figure 2. Dynamics in convalescent region: the red curve denotes the function $\mu_{t+1} = \mathcal{M}(\mu_t)$, blue arrows depict the equilibrium trajectory of μ_t.

[9] Formally, for any $T \in \mathbb{N}$, there exists $\varepsilon > 0$ such that if $\mu_0 \in (\widetilde{\mu} - \varepsilon, \widetilde{\mu})$, $\mu_t > 0$ for all $t < T$.

[10] Formally, for any $\delta > 0$, $T \in \mathbb{N}$, there exists $\varepsilon > 0$ such that if $\mu_0 \in (\widetilde{\mu} - \varepsilon, \widetilde{\mu})$, $\mu_t > \mu_0 - \delta$ for all $t < T$.

Figure 2 illustrates the gradual decline in μ_t, described in Proposition 2, by depicting the equilibrium starting from a point μ_0 in the convalescent region. The horizontal axis denotes μ_t, the vertical axis denotes μ_{t+1}, and the red curve denotes the function $\mu_{t+1} = \mathcal{M}(\mu_t)$. In the example depicted, μ_0 is shown to lie in the interval $I^5 = (\mu^4, \mu^5]$, so it takes 5 periods for the economy to reach the healthy region, and 7 periods to reach full employment. During the transition, employment is growing over time and the fraction of unskilled job-seekers is shrinking. As long as the economy is in the convalescent region, labor markets are slack and real wages are low, $\omega_t^* < \omega_{\text{fe}}^*$. But as soon as the economy reaches the interior of the healthy region, labor markets become tight and real wages reach their steady state level ω_{fe}^*. The real wage level $\omega^*(\mu_t)$, which we will refer to as the *natural real wage*, will play an important role in our analysis of monetary policy in Section 4.1.[11]

When the fraction of unskilled job-seekers μ_0 is higher than $\underline{\mu}$, expected training costs $\chi\mu_0$ are so high that firms are unable to recoup these costs if wages are high and expected to stay so. Slack labor markets must therefore persist for some time for wages to be low and for firms to still be willing to post vacancies at date 0. In other words, in equilibrium, the labor market must experience a *slow recovery*. In fact, the speed of the recovery decreases in the initial fraction of unskilled job-seekers. A higher μ_0 requires a lower job finding rate for wages to be driven down and firms to be induced to post vacancies. But such a low job finding rate in turn reduces the rate at which unskilled job-seekers are re-hired and regain skill. In the convalescent region, $\mu_t - \mu_{t+1}$ (the gap between the 45 degree line and the red line in Figure 2) is a decreasing function of μ_t: the worse the current state of the labor market, the slower it recovers. Accordingly, the economy can spend an arbitrary long time in the convalescent region before transitioning to the healthy region. When the economy starts deep in the convalescent region (μ_0 close to $\widetilde{\mu}$) the recovery takes disproportionately longer (point 1 of Proposition 2), and the fraction of unskilled job-seekers declines at a slower rate in the early stage of the recovery (point 2).

Stagnant region If the economy starts in the stagnant region, $\mu_0 \in [\widetilde{\mu}, 1]$, it never returns to full employment. When $\mu_0 \in [\widetilde{\mu}, 1]$, expected training costs are so high that they discourage firms from posting enough vacancies to bring the economy out of this region. Importantly, this is *not* because real wages are sticky. In the stagnant region, the high fraction of unskilled job-seekers, μ_t, is accompanied by low real wages which induce firms to post some vacancies. But such low real wages can only be sustained by low job finding rates, which in turn prevent unskilled workers from being hired and retrained in sufficient numbers for the economy to escape the stagnant region. This region is an *unemployment trap*, as the following proposition states.

Proposition 3 (Unemployment traps). *If the economy is pushed into the stagnant region, i.e., $\mu_t \geq \widetilde{\mu}$, then it never returns to the full employment steady state.*

Proof. See Appendix E. □

[11] Sterk (2016) also presents a search model which features multiple steady states and slow recoveries. However, the mechanisms which drive such outcomes in his model are different than in our model. Sterk (2016) assumes that the workers have bargaining power $\eta = 0$, implying that wages are always equal to b regardless of the fraction of unskilled workers. Thus, the *natural real wage* in his model is constant. In contrast, in our model, a time-varying natural wage is an important aspect of a slow recovery, as explained above.

The area to the right of $\widetilde{\mu}$ in Figure 1 depicts the stagnant region. Starting from any μ_0, there is a unique equilibrium in which μ_t converges to the high unemployment steady state $\overline{\mu}$.[12] Thus, if the economy starts in the stagnant region, it never escapes. In other words, the model features multiple steady states but a unique equilibrium. We discuss this in greater detail in Section 5.3.

4 Nominal rigidities

The analysis above highlighted that starting from a high level of unemployment, the economy may be unable to return to full employment. With nominal rigidities, this means that if monetary policy fails to act quickly and prevent large adverse shocks from significantly increasing unemployment, such shocks can have persistent or permanent consequences.

Shocks We focus on the economy's response to a temporary demand shock, modeled as a temporary increase in households' patience: $\beta_0 > 1$, $\beta_t = \beta < 1$ for all $t > 0$. The NK literature has used this type of shock to capture an increase in the supply of savings which pushes the real interest rate below zero.[13] We prefer to focus on a temporary demand shock (rather than, e.g., a productivity shock) since such a shock can only have persistent effects in the presence of nominal rigidities.

Nominal rigidities The model specified in the previous section is characterized by the classical dichotomy and thus, monetary policy is unable to affect allocations. Since our objective is to understand whether monetary policy can prevent or moderate hysteresis, we break the classical dichotomy by introducing nominal rigidities in the form of downwardly sticky nominal wages. In particular, we suppose that at any date t the nominal wage must satisfy $W_t \geq \varphi W_{t-1}$ where $\varphi \in (0, 1]$ limits how much nominal wages can fall between dates $t-1$ and t ($\varphi = 1$ means that nominal wages cannot fall, while $\varphi < 1$ implies that nominal wages can adjust downwards to some extent).

In the spirit of Schmitt-Grohe and Uribe (2013), given the current state μ_t, we assume that nominal wages are set to $W_t = \omega^*(\mu_t)P_t$ whenever possible, where $\omega^*(\mu_t)$ is the real wage in the flexible wage benchmark. However, if $\omega^*(\mu_t)P_t < \varphi W_{t-1}$, then $W_t = \varphi W_{t-1}$. That is, the nominal wage is set by Nash bargaining whenever the downward nominal wage rigidity (DNWR) constraint is not violated:

$$W_t = \max\Big\{\varphi W_{t-1}, P_t \omega^*(\mu_t)\Big\}. \tag{12}$$

Note that even when the DNWR constraint binds and nominal wages are unable to further adjust

[12] Appendix E proves that the economy never leaves the stagnant region. To show that there is unique equilibrium which converges to $\overline{\mu}$, we use numerical methods. We check the stability properties of the high-unemployment steady state by drawing a random sample of a 100,000 parameter combinations, dropping any draw which violates Assumptions 1 and 2. Given a parameter vector, we then linearize our two dynamic equations around the high-unemployment steady state and compute the eigenvalues of this system. Since we have one pre-determined variable (μ_t) and one jump-variable (θ_t), a unique equilibrium converging to $\overline{\mu}$ exists if one eigenvalue is inside the unit circle and one is outside. 100 percent of our feasible draws satisfy this property, indicating that equilibrium is unique in the stagnant region.

[13] In a richer model, such a shock could arise from a tightening of borrowing limits or an increase in precautionary savings motives. See, for example, Eggertsson and Krugman (2012) and Guerrieri and Lorenzoni (2017).

downwards, *real* wages can still fall if inflation is positive. Indeed, the real wage is given by

$$\frac{W_t}{P_t} = \omega_t = \max\left\{\varphi\frac{P_{t-1}}{P_t}\omega_{t-1}, \omega^*(\mu_t)\right\}. \tag{13}$$

Our choice of DNWR rather than symmetric wage rigidity is motivated by the overwhelming empirical evidence supporting asymmetric wage rigidities, e.g., Grigsby, Hurst and Yildirmaz (2019), Hazell and Taska (2019), Fallick et al. (2020) and other papers cited in our literature review. These papers present evidence that nominal wages are downwardly sticky for all workers, including new hires.

It is important to note that our results do not require nominal wages of new and existing workers to be *very* downwardly inflexible; all we need is that nominal wages are not perfectly flexible. Furthermore, in all the scenarios we consider, the real wage stays within the bargaining set, and so the workers have no incentive to agree to a further nominal wage cut (beyond the fall from W_{t-1} to φW_{t-1}). If instead, shocks were larger than in the scenarios we consider, leading real wages to leave the bargaining set under specification (12), we could replace (12) with $W_t = \min\left[\max\left[\varphi W_{t-1}, P_t\omega^*\left(\mu_t\right)\right], P_t\overline{\omega}_t\right]$ where $\overline{\omega}_t = A + \beta\left(1-\delta\right)J_{t+1}$ is the highest wage which lies in the bargaining set. This would imply that firms and workers renegotiate whenever nominal rigidities drive the wage outside the bargaining set. Making this assumption would not change any of our results. In this sense, we do not require that workers and firms forgo mutually beneficial wage cuts and the Barro (1977) critique does not apply.

4.1 Response under Conventional Monetary Policy (CMP)

We first examine how the economy would respond to a temporary demand shock under a regime we term *conventional monetary policy* (CMP).

Description of CMP We assume that the monetary authority sets the nominal interest rate i_t subject to the ZLB, i.e., $i_t \geq 0$. Risk neutrality implies that the equilibrium real interest rate equals $r_t = \beta_t^{-1} - 1$. Inflation then follows from the Fisher equation:

$$\frac{P_{t+1}}{P_t} = \beta_t(1 + i_t). \tag{14}$$

Since nominal wages are not perfectly flexible downwards, monetary policy can affect real wages and the level of unemployment by influencing the price level. We specify conventional monetary policy as a standard interest rate rule responding, whenever possible, to deviations of the price level from a state-dependent price level target $P_t^\star$ compatible with the real wage being at its natural level for an unchanged nominal wage:

$$1 + i_t = \max\left\{\beta_t^{-1}\left(\frac{P_t}{P_t^\star}\right)^{\phi_p}, 1\right\} \qquad \text{where} \qquad P_t^\star = \frac{W_{t-1}}{\omega^*(\mu_t)}. \tag{15}$$

It is worth noting that implementing the price level $P_t = P_t^\star$ results in the lowest possible unemployment rate given the state of the economy, while keeping nominal wage inflation at zero. To see this, note first that dividing (12) by P_t and W_{t-1} respectively yields:

$$\frac{W_t}{P_t} = \max\left\{\varphi\frac{W_{t-1}}{P_t}, \omega^*(\mu_t)\right\} \qquad \text{and} \qquad \frac{W_t}{W_{t-1}} = \max\left\{\varphi, \frac{P_t\omega^*(\mu_t)}{W_{t-1}}\right\}.$$

When $P_t < \varphi P_t^\star$, the DNWR constraint binds and raising P_t reduces real wages (encouraging hiring, reducing unemployment) without any effect on nominal wage inflation. On the other hand, when $P_t \geq \varphi P_t^\star$, the DNWR does not bind. In that case, an increase in P_t has no effect on real wages (and thus on unemployment) but translates one-for-one into higher nominal wage inflation. Thus, the price level $P_t = P_t^\star$ implements the lowest possible level of real wages (and unemployment) while keeping nominal wage inflation at zero. Our labelling of this policy as "conventional" reflects its similarity to interest rate rules studied in the NK literature, as well as its closeness to the US Federal Reserve's policy framework.[14]

The interest rate rule (15) indicates that monetary policy raises rates whenever prices exceed their target $P_t^\star$, and cuts rates when prices undershoot their target, subject to the ZLB. It is convenient to assume a very strong monetary response, i.e., $\phi_p \to \infty$ in equation (15), implying that in equilibrium:

$$P_t \leq \frac{W_{t-1}}{\omega^*(\mu_t)}, \qquad i_t \geq 0, \qquad \text{with at least one equality.} \tag{16}$$

In words, either the ZLB does not bind and prices are at their target level, or the ZLB binds and prices are below target. Importantly, as shown in Appendix F, CMP implements the same allocations as would be chosen by a monetary policymaker acting optimally under discretion.[15]

Given initial conditions μ_0 and W_{-1}, the equilibrium under CMP is a sequence $\{W_t, P_t, i_t, J_t, \omega_t, \theta_t, \mu_{t+1}\}_{t=0}^{\infty}$ satisfying (6), (7), (10), (12), (14), (16) and $\omega_t = W_t/P_t$ for all $t \geq 0$. To understand the economy's response to a demand shock, it is useful to note that given the wage setting rule (12) and the policy rule, nominal wages never increase in equilibrium even though they are fully flexible upwards. For ease of exposition, we focus below on the case with $\varphi = 1$ (nominal wages cannot fall).[16] In that case, equilibrium nominal wages are constant ($W_t = W_{-1}$ for all t), so the path of the real wage directly reflects that of the price level.

Effects of a transitory discount factor shock Assuming that the economy is initially at the full employment steady state ($\mu_0 = 0$), we now describe its response to a transitory demand shock ($\beta_0 > 1$, $\beta_t = \beta < 1$ for all $t > 0$) under CMP. This shock causes the ZLB to bind and causes the date 0 price level to fall. Since nominal wages cannot fall, real wages increase, discouraging firms

[14] Like the monetary authority in our model which aims to implement zero wage inflation and to keep unemployment at its "natural" flexible wage level, the Fed has had targets for both prices and real activity. In terms of prices, the Fed targets price inflation rather than nominal wage inflation; in our model, it makes more sense for the monetary authority to target nominal wages since these are the prices that are sticky (Aoki, 2001).

[15] Appendix F shows that this policy is optimal under discretion for a planner who minimizes $(u_t - 0)^2 + \lambda\left(\frac{W_t}{W_{t-1}} - 1\right)^2$, where λ is the relative weight the planner puts on stabilizing nominal wage inflation. CMP is the solution to the discretionary planner's problem for any $\lambda > 0$, however small.

[16] At the end of this section, we show that our characterization of the equilibrium dynamics under CMP is unaffected by this restriction.

from posting vacancies. For small shocks, while vacancy creation falls, the economy remains at full employment. Larger shocks cause unemployment to rise; since unemployment erodes human capital, this increases the fraction of unskilled job seekers μ_1, driving the economy into the convalescent region and resulting in a slow recovery. If the shocks are larger still, μ_1 may enter the stagnant region, resulting in permanent stagnation. These dynamics are summarized in Proposition 4.

Proposition 4 (Conventional monetary policy)**.** *Assuming that the economy is initially at the full-employment steady state ($\mu_0 = 0$), there exists a $\underline{\beta} = \frac{A-\kappa}{\omega^*_{\text{fe}}-(1-\delta)J_{min}} > 1$ such that:*

1. *If the shock is not too large, $\beta_0 \in (1, \underline{\beta}]$, vacancy posting falls, $\theta_0 \in (1, \theta^{\text{fe}})$, but the economy remains at full employment, $\mu_t = 0$ for all t.*

2. *For a large enough shock, $\beta_0 > \underline{\beta}$, labor markets are slack, $\theta_0 \in [0, 1)$, and unemployment rises, $\mu_1 \in (\underline{\mu}, \mu_R]$, where $\mu_R = (2-\delta)^{-1}$ is the rate of skill depreciation after a one-period hiring freeze, starting from full employment.*

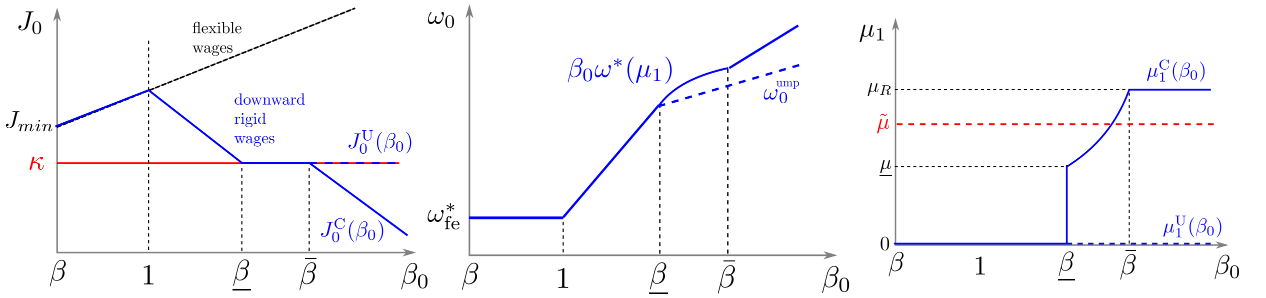
Furthermore, for $\beta_0 > \underline{\beta}$, if $\mu_1 < \widetilde{\mu}$, then the economy eventually returns to full employment whereas if $\mu_1 \geq \widetilde{\mu}$, then the economy never returns to full employment.

Proof. See Appendix H. □

In the flexible wage benchmark, a temporary increase in β_0 would not raise unemployment. In fact, since a filled vacancy is a long-lived asset yielding dividends in the future and the cost of posting a vacancy is paid today, a temporary increase in the discount factor increases the net present value of vacancy posting, *encouraging* vacancy creation (a *neoclassical effect*). However, with nominal rigidities, under CMP, outcomes differ from the flexible wage benchmark economy when the ZLB binds. To be clear, this does not mean that it is impossible for *any* monetary policy to replicate the flexible wage outcome when the ZLB binds - in the next section, we will study an unconventional monetary policy that does exactly that. It only means that *conventional* monetary policy fails to do this.

All else equal, a higher β_0 increases households' demand for bonds which lowers the demand for current consumption and puts downward pressure on its price P_0. CMP tries to prevent P_0 from falling by lowering the nominal rate, dissipating the excess demand for bonds. When the ZLB binds, the nominal return on bonds cannot be lowered any further and the Fisher equation (14) indicates that inflation between dates 0 and 1 must satisfy $P_1/P_0 = \beta_0 > 1$. Whether this requirement is met in equilibrium by a rise in P_1 or a fall in P_0 depends on the conduct of monetary policy.

Under CMP, P_1 never rises enough to prevent a fall in P_0 in equilibrium: P_0 must fall in response to a demand shock of magnitude $\beta_0 > 1$. Accordingly, real wages and the cost of hiring at date 0 rise with the decline in P_0. To see why P_0 must fall, suppose that it instead remains constant at P_{-1}. Since the ZLB binds at date 0, this implies that P_1 must rise to $P_1 = \beta_0 P_{-1}$. Accordingly, real wages remain at ω^*_{fe} at date 0 and 1, implying that nominal wages increase between date 0 and 1, $W_1/W_0 > 1$. But since the ZLB does not bind at date 1, CMP implements $P_1^\star = W_0/\omega^*_{\text{fe}}$ (see equation (16)). This, however, implies that $W_1/W_0 = 1$, which is a contradiction. Thus, P_0 must fall. Intuitively, because CMP aims to implement a price level consistent with zero nominal wage inflation at date 1, P_1 cannot rise enough to prevent P_0 from falling in equilibrium.



(a) Value of a vacancy **(b)** Date 0 real wages **(c)** Skill composition of job-seekers

Figure 3. Response to a demand shock: Panel (a) - solid blue curve depicts the relationship between J_0 and β_0 under CMP, blue-dashed line represents the same relationship under UMP, dashed-black line represents the same relationship absent nominal rigidities, red line represents the cost of job-creation at date 0. Panel(b) - solid blue curve depicts the relationship between ω_0 and β_0 under CMP, dashed-blue curve represents same relationship under UMP. Panel (c) - solid blue curve represents the relationship between μ_1 and β_0 under CMP, dashed-blue line represents the same relationship under UMP.

Since nominal wages cannot fall, a lower P_0 raises real wages at date 0. Thus, when the ZLB binds, the neoclassical effect of a higher β_0 can be outweighed by a deflationary *Keynesian effect*, causing a fall in the value of vacancy creation. With perfectly flexible nominal wages, only the neoclassical effect would operate, generating a positive relationship between the value of a filled vacancy J_0 and size of the shock β_0, as shown by the dashed upward sloping line in Figure 3a. In contrast, with nominal rigidities, the Keynesian effect is also at work, and under Assumption 2, it always dominates the neoclassical effect, resulting in a negative relationship between J_0 and β_0 whenever $\beta_0 > 1$, as shown by the solid downward sloping line in Figure 3a.

As stated in Proposition 4, the ultimate effect of a higher β_0 on the labor market depends on the size of the shock. Figures 3a, 3b and 3c depict the date 0 value of a filled vacancy, the date 0 real wage, and the date 1 fraction of unskilled job seekers, respectively, as functions of β_0. We next turn to a detailed dicussion of these relationships.

Moderate shocks When the increase in β_0 is moderate $\beta_0 \in (1, \underline{\beta}]$, while real wages increase and the value of a filled vacancy falls, this value does not fall enough to reduce vacancy posting substantially. Consequently, the economy remains at full employment ($\mu_1 = 0$). This means that under CMP, $P_1 = W_{-1}/\omega^*_{\mathrm{fe}}$. The binding ZLB at date 0 then implies that P_0 falls to $P_1/\beta_0 < P_{-1}$, and the real wage rises to $\omega_0 = \beta_0\omega^*_{\mathrm{fe}} > \omega^*_{\mathrm{fe}}$ (see Figure 3b). This reduces the value of a filled vacancy:

$$J_0 = \underbrace{A - \beta_0\omega^*_{\mathrm{fe}}}_{\text{current profit}} + \underbrace{\beta_0(1-\delta)J_{min}}_{\text{valuation of future profits}}. \tag{17}$$

The two opposite effects of a higher β_0 onto the value of a filled vacancy are visible in (17): a higher real wage lowers current profits (Keynesian effect) but a higher β_0 increases the valuation of future profits (neoclassical effect). Appendix G shows that under Assumption 2, the Keynesian effect dominates, i.e., $\partial J_0/\partial\beta_0 = -\omega^*_{\mathrm{fe}} + (1-\delta)J_{min} < 0$. For $\beta \in (1, \underline{\beta}]$, the value of a filled vacancy J_0 falls with β_0, but remains above the vacancy posting cost κ (see Figure 3a), implying that firms are still willing to post enough vacancies to keep the economy at full employment ($\mu_1 = 0$), as shown in Figure 3c.

Large shocks When the shock is larger, $\beta_0 > \underline{\beta}$, the real wage at date 0 increases by so much that the value of a filled vacancy either falls to or below the vacancy posting cost κ. Firms are not willing to post vacancies if they expect full employment (and thus high wages) to prevail in the future. Unemployment must thus rise, along with a worsening in the skill composition, for future wages to be low enough to possibly convince firms to keep hiring at date 0. To see this, note that the date 0 value of a filled vacancy is now given by

$$J_0 = \underbrace{A - \beta_0\omega^*(\mu_1)}_{\text{current profit}} + \underbrace{\beta_0(1-\delta)(\kappa + \chi\mu_1)}_{\text{valuation of future profits}} \quad \text{where } \mu_1 \in [\underline{\mu}, \mu_R]. \tag{18}$$

For firms to post vacancies at date 0 despite a high β_0, future profits must rise just enough to make the value of a filled vacancy J_0 in (18) at least as large as the vacancy posting cost κ. For future profits to rise, the fraction of unskilled job seekers must be high enough to generate low real wages even after the shock has abated: the economy enters the convalescent or stagnant region, i.e., $\mu_1 > \underline{\mu}$ (see Figure 3c). This rise in μ_1 means that CMP targets a higher price level $P_1^\star = W_0/\omega^*(\mu_1)$, which implies that P_0 falls by less from the Fisher equation. This mitigates but does not wholly prevent the rise in date 0 real wages, as shown by the solid blue curve between $\underline{\beta}$ and $\overline{\beta}$ in Figure 3b.

For $\beta_0 \in (\underline{\beta}, \overline{\beta}]$, the rise in μ_1 increases future profits enough to offset the decline in current profits, leaving $J_0 = \kappa$ in this region (flat region of the solid blue curve between $\underline{\beta}$ and $\overline{\beta}$ in Figure 3a). However, for extremely large demand shocks ($\beta_0 > \overline{\beta}$) the value of the firm J_0 falls below κ, and firms are no longer willing to post new vacancies - the hiring rate falls to zero and the date 1 fraction of unskilled job seekers increases to μ_R.[17] While this rise in μ_1 increases future profits, it is not enough to offset the fall in current profits and prevent J_0 from falling below κ. Note that while firms are unwilling to post new vacancies when $J_0 < \kappa$, they do not wish to destroy existing jobs, which also pay higher real wages, as long as $J_0 > 0$. This is true in all the scenarios we consider. That is, nominal rigidities are never so severe that they drive the real wage out of the bargaining set. Since existing matches still observe positive surplus, firms have no incentive to shut down and employed workers have no incentive to agree to wage cuts. Workers and firms do not forgo mutually beneficial wage cuts and the Barro (1977) critique does not apply.

In the event of a large increase in unemployment, the economy experiences either a slow recovery or a permanent stagnation. Which of these scenarios realizes depends both on the size of the shock and on the forces generating multiple steady states. The higher the training cost χ, the lower the threshold $\widetilde{\mu}$, and the more likely it is that the economy enters the stagnant region following a sufficiently large shock.[18] Similarly, for a given χ, a larger β_0 is more likely to push the economy into the stagnant

[17] See Appendix H for the definition of $\overline{\beta}$. In our economy with only one period shock, μ_R is the maximum damage that can be inflicted on the skill composition of the workforce during the ZLB episode, starting from full employment. Shocks which last longer could of course result in a higher μ.

[18] More generally, shocks lasting multiple periods would also be more likely to bring the economy to the stagnant region. In this section we focus on one period shocks to emphasize that even very transitory recessions can have permanent effects. However, in our quantitative analysis in Appendix L, we allow for persistent shocks.

region. Figure 3c depicts the case in which $\mu_R > \widetilde{\mu}$, so a large enough shock that causes a hiring freeze at date 0 always drives the economy to the stagnant region.

Slow recovery When the reduction in hiring drives the economy into the convalescent region, i.e., when $\mu_1 \in (\underline{\mu}, \widetilde{\mu})$, the economy ultimately returns to full employment, but the recovery takes time. The dashed blue line in Figure 4b depicts the dynamics of μ following a shock which drives the economy to the convalescent region. Following an initial deterioration in the skill composition of the workforce due to a hiring slump, the economic forces underlying this slow recovery are essentially the ones outlined in Section 3. Faced with a higher likelihood of meeting unskilled applicants and hence higher expected training costs, firms only post vacancies if they are compensated by lower real wages. In turn, the only way low wages can be an equilibrium outcome is if job-finding rates are depressed for some time, keeping the worker's outside option low. As a result, the unemployment rate and the fraction of unskilled job-seekers only decline gradually, but the economy ultimately returns to full employment.

Permanent stagnation When the date 0 hiring slump takes the economy into the stagnant region ($\mu_1 \geq \widetilde{\mu}$), the economy never returns to full employment. The solid red line in Figure 4b shows the dynamics of μ in this case. Again, conditional on the initial deterioration of the skill composition, the forces behind the ensuing stagnation dynamics are not nominal but real – the DNWR constraint does not bind beyond date 0. Unemployment remains permanently high *not* because of nominal frictions but because of a deterioration of unemployed workers' human capital. The fraction of unskilled job-seekers is so high that real wages must be very low for firms to post *any* vacancies. Such low real wages can only be sustained if slack labor markets are expected to persist forever, implying that the economy converges to the high unemployment steady state. In this steady state, even though high unemployment depresses wages, firms are reluctant to post vacancies because the average job-seeker is likely to be unskilled and costly to retrain. These low vacancy posting rates support high unemployment. Thus, even a transitory demand shock can permanently depress employment.

Dynamics of prices, wages and unemployment After date 1, the evolution of the price level depends on whether the economy eventually returns to full employment, or converges to the high unemployment steady state. If the economy returns to full employment, the real wage eventually returns to ω^*_{fe} (as shown in Figure 4c), and the (targeted and realized) price level falls back to P_{-1} (shown by the dashed blue line in Figure 4a). If instead the economy converges to the high unemployment steady state, the real wage falls further to $\omega^*(\overline{\mu})$, and the price level rises further to $W_{-1}/\omega^*(\overline{\mu})$ (solid red line in Figure 4a). Regardless of the scenario, the price level rises above its pre-shock value at date 1 when $\beta_0 > \underline{\beta}$, but does not rise enough to prevent unemployment at date 0 (as shown in Figure 4d). The unconventional monetary policy we propose in Section 4.2 does not share this shortcoming.

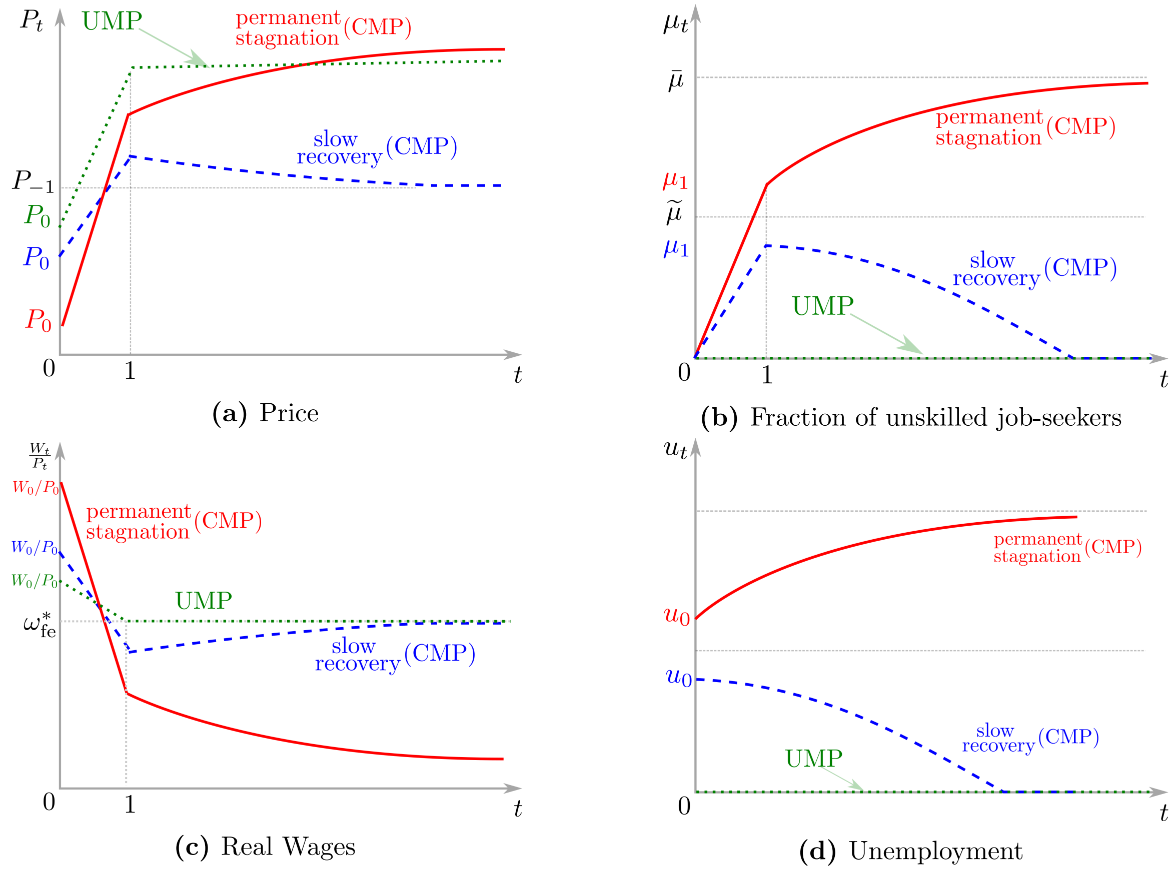


Figure 4. Dynamics under CMP and UMP: Panel (a) depicts the trajectory of the price level, panel (b) depicts the trajectory of μ, panel (c) depicts trajectory of real wages and panel (d) depicts the trajectory of the unemployment rate. In all panels, the red solid curves depict the case where the economy features permanent stagnation under CMP, the dashed-blue lines depict the case where the economy features a slow recovery under CMP and the green dotted lines depict the case under UMP.

4.1.1 Case with $\varphi < 1$

Before discussing how unconventional monetary policy can prevent hysteresis from taking root, it is useful to highlight that our results do not depend on nominal wages being fully rigid downward. The discussion above assumed $\varphi = 1$, i.e., nominal wages cannot fall however large the shock or level of unemployment. Nonetheless, Proposition 4 remains true even if we allow nominal wages to fall by some amount $(0 < \varphi < 1)$ and even if this degree of flexibility is state dependent, as long as nominal wages are not fully flexible $(\varphi = 0)$. In particular, the dynamics of all real variables remain the same as described above even if nominal wages are allowed to fall somewhat.

With $\varphi < 1$, following a shock $\beta_0 > 1$, nominal wages do fall by the maximum amount possible to $W_0 = \varphi W_{-1}$. While in partial equilibrium, such a fall in nominal wages would encourage firms to post vacancies, this increased incentive to post vacancies is negated by general equilibrium forces – prices fall by a larger amount, leading real wages to rise to the exact same level as when $\varphi = 1$. To see why, recall that since the ZLB does not bind at date 1, the CMP implements $P_1 = P_1^\star = W_0/\omega^*(\mu_1)$.

However, the ZLB does bind at date 0, so the Euler equation implies

$$P_0 = \frac{1}{\beta_0} P_1 = \frac{1}{\beta_0} \frac{W_0}{\omega^* (\mu_1)}.$$

Thus, the real wage satisfies $\omega_0 = \beta_0 \omega^* (\mu_1)$ and the date 0 value of a filled vacancy is given by $J_0 = A - \beta_0 \omega^* (\mu_1) + \beta_0 (1 - \delta) J_1$, exactly as in the $\varphi = 1$ case. And since the ZLB and DNWR constraints do not bind after 0, all remaining equations characterizing the dynamics of the real variables are identical to the ones obtained when $\varphi = 1$, and the proof of Proposition 4 equally applies (see Appendix J). Thus, allowing the date 0 nominal wage W_0 to adjust downwards by some degree triggers a proportional decline in the date 0 price level P_0, resulting in the real wage ω_0 remaining the same as when $\varphi = 1$. Therefore, greater nominal wage flexibility does not mitigate the rise in real wages.

Intuitively, CMP anchors all future nominal variables to the date 0 nominal wage. Therefore, any additional drop in the date 0 nominal wage is accompanied by a fall of the same magnitude in nominal variables at date 1 and beyond. But by targeting a lower price level P_1 at date 1, the monetary authority further exacerbates the decline in households' consumption demand at date 0. Accordingly, date 0 prices must fall even further so as to dissipate excess demand on the date 0 bond market, leaving real wages unchanged relative to the case where $\varphi = 1$.

Since this argument was made for any $\varphi \in (0, 1)$, it also explains why our results remain unchanged if the degree of nominal wage flexibility φ is state-dependent, as Appendix J shows.

4.2 Response under Unconventional Monetary Policy (UMP)

CMP arguably provides a good description of the conduct of monetary policy in many advanced economies (at least prior to the Great Recession), and – as already mentioned – it is optimal under discretion (see Appendix F). However, as we have shown, this common approach to policy can lead to undesirable outcomes in the presence of hysteresis.

The problem with CMP is that when the ZLB binds in the present, policy fails to commit to more expansionary policy in the future to forestall deflation today. As we now show, an *unconventional monetary policy* (UMP) which makes such commitments mitigates deflationary forces at date 0 and prevents the initial rise in unemployment and the subsequent damage, even for large shocks. Similar to CMP, we specify UMP as attempting to implement a time-varying price level target $\{\widetilde{P}_t^\star\}_{t=0}^\infty$. In particular, when the initial shock β_0 is small ($\beta_0 \leq \underline{\beta}$), then the price level target $\widetilde{P}_t^\star$ under UMP is identical to the target $P_t^\star$ under CMP. However, when the shock is larger than $\underline{\beta}$, UMP is more accommodative and commits to a higher price level target at date 1 ($\widetilde{P}_1^\star > P_1^\star$) and beyond. Specifically, for $\beta_0 > \underline{\beta}$, $\{\widetilde{P}_t^\star\}_{t=0}^\infty$ is given by:

$$\widetilde{P}_0^\star \quad = \quad \frac{W_{-1}}{\omega_0^{\text{ump}}(\beta_0)} \qquad \text{and} \qquad \widetilde{P}_t^\star = \beta_0 \widetilde{P}_0^\star \qquad \text{for all } t \geq 1 \tag{19}$$

where $\omega_0^{\text{ump}}(\beta_0) = A - \kappa + \beta_0 (1 - \delta) J_{min}$ is the real wage which sets the value of a filled vacancy at date 0 to κ ($J_0 = \kappa$). Note that since $\omega_0^{\text{ump}}(\beta_0) > \omega_{\text{fe}}^*$, we have $\widetilde{P}_0^\star < P_{-1} = W_{-1}/\omega_{\text{fe}}^*$, so this policy permits some deflation at date 0 – but crucially, *less* deflation than observed in equilibrium under

CMP. Proposition 5 describes how outcomes differ under UMP when the economy experiences a shock β_0 large enough that it would move away from full employment under CMP.

Proposition 5 (Unconventional monetary policy)**.** *Suppose that $\beta_0 > \underline{\beta}$ and the economy starts at full employment, i.e., $\mu_0 = 0$ and $W_{-1} = \omega^*_{\text{fe}} P_{-1}$. Under UMP, the economy remains at full employment, $\mu_t = 0$ for all $t \geq 0$. Prices are equal to target at all dates $t \geq 0$, $P_t = \widetilde{P}^\star_t$ for $\widetilde{P}^\star_t$ described in (19). Nominal wages are given by $W_0 = W_{-1}$ and $W_t = \beta_0 \frac{\omega^*_{\text{fe}}}{\omega^{ump}_0(\beta_0)} W_{-1} > W_0$ for all $t \geq 1$. Real wages are given by $\omega_0 = \omega^{ump}_0(\beta_0)$ and $\omega_t = \omega^*_{\text{fe}}$ for all $t \geq 1$.*

Proof. See Appendix I □

Recall that under CMP, the ZLB binds at date 0 and the price level P_0 falls to ensure that $P_1/P_0 = \beta_0 > 1$. While the future price level P_1 does rise under CMP, it does not rise enough to prevent a large fall in current prices P_0, which raises real wages and reduces hiring. In contrast, UMP commits to implement a date 1 price level high enough to prevent date 0 prices from falling to a level which would reduce hiring. While a large β_0 shock increases the demand for savings, a higher future price level P_1 attenuates this increase by making future consumption more expensive. The resulting higher demand for date 0 consumption hence mitigates the fall in P_0. Thus, while the date 0 real wage does rise under UMP, it only rises to $\omega^{\text{ump}}_0(\beta_0)$ (shown by the dashed blue line in Figure 3b). As a result, the value of a filled vacancy J_0 never falls below the cost of posting a vacancy κ, however large β_0 is (dashed blue line in Figure 3a). Firms are willing to post enough vacancies to keep the economy at full employment despite the fall in prices. And since μ_1 does not rises above 0 (dashed blue in Figure 3c), the economy never enters the convalescent or stagnant regions and hysteresis is averted. However, this policy comes at the cost of positive nominal wage inflation between dates 0 and 1. The time paths of the price level, the fraction of unskilled job-seekers, the real wage and rate of unemployment are represented by the dotted green lines in Figures 4a, 4b, 4c and 4d, respectively.

UMP is similar to *forward guidance* policies (Eggertsson and Woodford, 2003; Werning, 2011) in that when the ZLB binds, it commits to more accommodative policy in the future. However, while forward guidance is usually perceived as a commitment to lower nominal rates in the future, UMP involves a commitment to target higher prices at date 1, but does not necessarily involve a lower path of nominal rates compared to CMP. Section 4.1 described how CMP can be thought of as a limiting case of a Taylor type rule (15) which raises rates aggressively when prices exceed their target, and cuts rates aggressively when prices undershoot their target, until nominal rates hit the ZLB. UMP can be thought of in the same way, but with a different price level target. Following a large shock at date 0, UMP targets a higher date 1 price level than CMP ($\widetilde{P}^\star_1 > P^\star_1$). Off-equilibrium, UMP would cut interest rates more aggressively than CMP at date 1, if faced with the same observed price level P_1. On equilibrium, however, prices do not actually deviate from target at date 1 (since the ZLB does not bind at date 1) under either CMP or UMP. Thus, it is not necessarily the case that the unconventional policymaker implements lower nominal rates at date 1 in equilibrium.[19]

[19] In our economy with linear utility – and hence exogenous real interest rates, the nominal interest rate behaves analogously to expected inflation. In the slow recovery scenario, prices decline after date 1 under CMP (Figure 4a), so expected inflation and nominal rates are lower than under UMP, which features constant prices after date 1. In contrast,

5 Discussion

5.1 Commitment vs discretion and the importance of timely accommodation

The difference between CMP and UMP is not that the ZLB binds under one policy and not the other: the ZLB also binds under UMP at date 0. Instead, the key difference is that UMP makes commitment regarding future policy. Indeed, just as CMP is optimal under discretion for a planner who minimizes the loss function $u_t^2 + \lambda \left(\frac{W_t}{W_{t-1}} - 1 \right)^2$, UMP is optimal under commitment for a planner with the same preferences, provided that λ is sufficiently small (see Appendix F).[20] Thus, one interpretation of our results is that a discretionary policymaker cannot prevent adverse shocks from causing hysteresis. This is because the policy required to avoid hysteresis is a timely commitment to more accommodative policy in the future (date 0 commitment to a high enough P_1).

However, the ability to commit is by itself not enough to overcome hysteresis. It is essential that commitments be made in a *timely* fashion, before unemployment has increased and the skill composition of job-seekers has deteriorated. Suppose that monetary policy fails to make such commitments at date 0 and is expected to follow CMP at all future dates, allowing unemployment to rise and pushing μ_1 into the stagnant region. Can policy then reverse course and return the economy to full employment - e.g. by committing at date 1 to implement the price sequence $\{\widetilde{P}_t^\star\}_{t=1}^\infty$ from then onwards? The answer is no: monetary policy cannot engineer an escape from an unemployment trap. In the stagnant region, employers are unwilling to create more vacancies despite prevailing low real wages since the fraction of unskilled job seekers is large. If monetary policy could temporarily drive down real wages further, this would encourage hiring and bring down the unemployment rate. Recall, however, that our model features asymmetric nominal wage rigidities, in line with the empirical evidence. Nominal wages are rigid downwards, but flexible upwards: implementing a higher price level would only cause nominal wages to rise one-for-one, leaving real wages unchanged at their natural level. Similarly, in the event of a slow recovery rather than a permanent stagnation, monetary policy cannot speed up the recovery after date 0 if it was insufficiently accommodative to begin with: lower real wages would in principle stimulate hiring, but monetary policy cannot reduce real wages. This highlights that slow recoveries and permanent stagnation in our economy are *not* driven by nominal rigidities and a binding ZLB after date 0. They are driven by the elevated fraction of unskilled job seekers, a purely real factor. Only at date 0, when the DNWR binds and deflation drives real wages above their natural level, can monetary policy improve outcomes by mitigating the deflationary pressures. *Timely commitment*, rather than commitment per se, is needed to prevent hysteresis.

The point that commitment delivers better outcomes than discretion, particularly when the ZLB binds, is well known in the NK literature (see, e.g., Werning 2011). However, our emphasis on *timely* commitments is novel. While commitment to expansionary future policy is effective in standard NK models, it is equally effective at any point in the recession. Further, delaying monetary accommodation

in the permanent stagnation scenario, prices rise after date 1 under CMP, so nominal rates are higher under CMP than under UMP.

[20] While UMP involves some wage inflation at date 1, it avoids persistently or permanently higher unemployment, which outweighs the cost of inflation provided that the planner does not put too high a weight on stabilizing inflation.

in such models is costly, but the costs are only temporary. In contrast, delayed commitments are ineffective in our model, and a failure to make a timely commitment can have permanent costs if shocks are large.

5.2 Is timeliness generally important?

A timely response of monetary policy is of paramount importance in our baseline model, since if monetary policy fails to act at date 0, it is powerless to speed up a recovery or escape an unemployment trap at a later point. Two features of our baseline model drive this result. First, nominal wages are flexible upwards, implying that monetary policy cannot push real wages below their natural level by reducing prices to stimulate hiring after hysteresis has taken root. Second, we have abstracted from other policies, such as hiring and training subsidies, which could help bring the economy back to full employment. Thus, one might wonder whether timeliness would be as important if nominal wage rigidities were symmetric or if other fiscal policy instruments were available. As we show next, either symmetric wage rigidities or fiscal policy make it possible for policy to accelerate recoveries or escape traps ex post. However, doing so remains unattractive, as it involves higher costs than what would have been incurred under a timely monetary policy response preventing any increase in unemployment in the first place.

5.2.1 Symmetric nominal rigidities

Suppose that instead of being downwardly rigid (as suggested by the empirical evidence reviewed in our introduction), nominal wages were fixed at $\overline{W}$, arguably the simplest form of a symmetric rigidity. UMP at date 0 would again prevent any increase in unemployment. However, if an alternative policy allowed a large shock to push the economy into the convalescent or stagnant region, monetary policy would be able to speed up a recovery or escape from an unemployment trap at a later date, unlike under DNWR. With a fixed nominal wage, implementing persistently higher prices from date 1 onwards would lower real wages and encourage hiring even when the skill composition of job-seekers make expected training costs high. Such a strategy, however, entails larger losses than enacting UMP at date 0.

Specifically, since nominal wages are fixed in this case, we specify the planner's loss function in terms of unemployment and price (rather than wage) inflation: $u_t^2 + \lambda_p \left(\frac{P_t}{P_{t-1}} - 1\right)^2$. Allowing unemployment to rise at date 0 and acting to restore full employment at date 1 entails a higher loss than following UMP from date 0 onwards. Under UMP, there is no unemployment, a modest deflation at date 0 and a modest inflation at date 1, yielding a loss of

$$\mathcal{L}_0^{\text{ump}} = \lambda_p \left(\frac{\widetilde{P}_0^\star}{P_{-1}} - 1\right)^2 + \beta_0 \lambda_p \left(\frac{\widetilde{P}_1^\star}{\widetilde{P}_0^\star} - 1\right)^2 = \lambda_p \left(\frac{\omega_{\text{fe}}^*}{\omega_0^{\text{ump}}} - 1\right)^2 + \beta_0 \lambda_p \left(\beta_0 - 1\right)^2.$$

For concreteness, consider the following alternative policy: CMP is initially expected to apply from date 0 onwards, but the policymaker unexpectedly deviates from CMP at date 1 and instead implements a higher than expected price level so as to lower real wages and bring the economy back to full employment. At date 0, this policy acts exactly like CMP, thus clearly entailing higher deflation and unemployment than UMP (see Figure 4).[21] At date 1, while inflation equals $\beta_0 > 1$ under UMP (and CMP), it is higher than β_0 under the alternative policy, which implements a higher price than CMP from date 1 onwards to reduce real wages below $\omega^*(\mu)$ and bring the economy back to full employment. Thus, the alternative policy involves strictly higher losses than UMP, both at date 0 and at date 1.

In our baseline model with DNWR, a failure to act at date 0 forces the monetary authority to accept higher unemployment at date 1 — once μ is higher, no amount of inflation can bring unemployment back immediately (or perhaps ever). With symmetric rigidities, it remains true that a failure to act at date 0 leaves monetary policy facing a less favorable inflation-unemployment trade-off than would have obtained under UMP. With inflation at $P_1/P_0 = \beta_0 > 1$, unemployment is positive; reducing unemployment to zero is not impossible but requires inflation above β_0. Thus, even with symmetric rigidities, timeliness is important, not because it is impossible to reverse the shock's effects, but because doing so is more costly than committing early to higher future prices. While the data points to asymmetric rigidities as being a more empirically realistic way to model nominal wage rigidities, the logic of our timeliness argument holds more generally.

5.2.2 Fiscal policy

Our baseline model focuses on monetary rather than fiscal policy, reflecting the reality that fiscal policy is often imperfect and slow to respond to a downturn, leaving monetary policy to be a first responder when it comes to countercyclical stabilization. Enriching our baseline model with fiscal policy tools, such as hiring or training subsidies, could make the effects of hysteresis less severe. Indeed, compensating firms for each worker they train would be equivalent to lowering the private training cost χ, potentially speeding up a recovery or even lifting the economy out of the stagnant region. However, the fact that fiscal policy can make up for a failure of monetary policy to act early does not imply that fiscal policy is necessarily more appropriate to address hysteresis, nor does it makes timeliness less of a relevant issue. Even in the presence of fiscal policy, an appropriately designed monetary policy such as UMP is capable of keeping the economy at full employment and preventing any skill depreciation. While hiring and training subsidies might be effective at mitigating or reversing an increase in unemployment, they can hardly do better in terms of employment outcomes than a monetary policy which keeps the economy at full employment throughout.

In addition, fiscal policies are not costless. While UMP entails some nominal wage inflation, relying on ex post training subsidies also comes with costs: lower aggregate output (due to unemployment), lower consumption due to the real resources which must be spent on training, and distortionary costs of taxation to finance the subsidy. These costs may well be higher than the costs of temporarily higher inflation associated with UMP. Thus, even if fiscal policy can in principle be deployed to make up for an inadequate monetary policy response, it does not invalidate our main insight that timely monetary policy action is valuable in response to large adverse shocks when hysteresis forces are present. Finally, even if one were to design automatic stabilizers which boost hiring in recessions, they would only kick

[21] Figure 4 shows $\mu_1 > 0$, which, given the one-to-one mapping between μ_t and u_{t-1}, implies $u_0 > 0$.

in once unemployment has increased and would thus incur the same resource costs mentioned above. Thus, timely monetary policy action is still required to prevent hysteresis from taking root.

To reiterate, it is true in our model that once unemployment has reached a high level, it causes skill depreciation which monetary policy is ill-equipped to reverse. Such skill depreciation would be better addressed by labor-market specific policies such as hiring and training subsidies. But what monetary policy *can* do is prevent unemployment from rising in the first place, which avoids any skill depreciation, obviating the need for labor-market policies.

5.3 Comparison with recent papers studying permanent stagnation

In our baseline model, we highlighted how even a shock that lasts for one period can permanently move the economy away from full employment. As such, our analysis shares some similarities with a number of recent studies, such as Benigno and Fornaro (2017) and Schmitt-Grohe and Uribe (2017) which explore the possibility of permanent stagnation. However, there are two key differences that distinguish our analysis from theirs.

First, in our economy, a binding DNWR constraint at date 0 is necessary for demand shocks to move the economy away from full employment, but once unemployment has increased, nominal rigidities no longer bind and skill depreciation during unemployment – a purely real factor – causes slow recoveries or permanent stagnation. In fact, given the one period shock we consider, our results would remain the same even if we assumed that nominal wages were fully flexible after the first period. In contrast, in Benigno and Fornaro (2017) and Schmitt-Grohe and Uribe (2017) a shock which drives the economy to the ZLB can cause permanently higher unemployment, but only if the ZLB and DNWR also bind forever. While monetary policy and demand shocks can have permanent effects on unemployment in our economy, this is not because of a long run Phillips curve through which permanent deflation causes permanently higher unemployment (Benigno and Ricci, 2011) – our DNWR does not bind in steady state. Instead, temporary deflation can generate permanently higher unemployment.

Second, our paper brings the idea of path dependence into the literature on liquidity traps and secular stagnation. Again, this yields starkly different predictions from Benigno and Fornaro (2017) and Schmitt-Grohe and Uribe (2017). These papers feature multiple equilibria, one of which features deflation and high unemployment.[22] In these models, persistent unemployment can be an equilibrium outcome because agents' pessimistic beliefs are self-reinforcing. If agents, however, awoke one morning and expected the economy to return to full employment, the economy would indeed return to full employment. In contrast, our economy is not trapped in the high unemployment steady state because of self-fulfilling beliefs. In fact, starting from this steady state, if (off equilibrium) firms anticipated a return to full employment, they would be less willing to hire workers today, since they would anticipate a more skilled workforce and lower costs of job creation tomorrow. Lack of hiring today would further cement the skill deterioration in the workforce and reinforce high unemployment rates. Thus, persistently high unemployment arises in our model not because of self-fulfilling beliefs, but

[22] Of course, these models also feature multiple steady states in the sense that the economy can stay in the bad or good equilibrium forever.

because our economy features an endogenous, slow-moving state variable – the skill composition of job-seekers.

6 Conclusion

We presented a model designed to study the positive and normative implications of hysteresis. Skill depreciation, nominal rigidities and constraints on monetary policy together allow temporary shocks to generate slow recoveries or even permanent stagnation. Aggressive countercyclical policy may be able to avoid these outcomes, but only if enacted in a timely manner. While we have focused on skill depreciation, more generally recessions may damage productive capacity through multiple channels - reducing capital accumulation, reducing labor force participation, slowing productivity growth, and so on. Many of these effects may also be hard or even impossible to reverse. For example, Wee (2016) shows that recessions can permanently change young workers' search behavior, causing them to stay in careers in which they have a comparative disadvantage but have accumulated sufficient specific human capital, causing permanent misallocation. Whenever such mechanisms are operative, it is all the more important for countercyclical policy to nip recessions in the bud; the damage from failing to do so may be irreversible. In a world vulnerable to hysteresis, prevention is better than cure.

References

Acharya, Sushant, Edouard Challe, and Keshav Dogra, "Optimal Monetary Policy According to HANK," Staff Reports 916, Federal Reserve Bank of New York February 2020.

Aoki, Kosuke, "Optimal monetary policy responses to relative-price changes," *Journal of Monetary Economics*, 2001, *48* (1), 55 – 80.

Ball, Laurence, "Long-term damage from the Great Recession in OECD countries," *European Journal of Economics and Economic Policies: Intervention*, September 2014, *11* (2), 149–160.

Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk, "Some Evidence on the Importance of Sticky Wages," *American Economic Journal: Macroeconomics*, January 2014, *6* (1), 70–101.

Barro, Robert J., "Long-term contracting, sticky prices, and monetary policy," *Journal of Monetary Economics*, July 1977, *3* (3), 305–316.

Barron, John M, Dan A Black, and Mark A Loewenstein, "Job Matching and On-the-Job Training," *Journal of Labor Economics*, January 1989, *7* (1), 1–19.

Benigno, Gianluca and Luca Fornaro, "Stagnation Traps," *Review of Economic Studies*, 2017, *85* (3), 1425–1470.

Benigno, Pierpaolo and Luca Antonio Ricci, "The Inflation-Output Trade-Off with Downward Wage Rigidities," *American Economic Review*, June 2011, *101* (4), 1436–66.

Berger, David, Ian Dew-Becker, Konstantin Milbradt, Lawrence Schmidt, and Yuta Takahashi, "Layoff risk, the welfare cost of business cycles, and monetary policy," Technical Report 2016.

Bianchi, Francesco, Howard Kung, and Gonzalo Morales, "Growth, Slowdowns, and Recoveries," Working Paper 20725, National Bureau of Economic Research December 2014.

Caballero, Ricardo J and Emmanuel Farhi, "The Safety Trap," *Review of Economic Studies*, 2017, *85* (1), 223–274.

Correia, Isabel, Juan Pablo Nicolini, and Pedro Teles, "Optimal Fiscal and Monetary Policy: Equivalence Results," *Journal of Political Economy*, February 2008, *116* (1), 141–170.

Dickens, William T., "The productivity crisis: Secular or cyclical?," *Economics Letters*, 1982, *9* (1), 37 – 42.

Drazen, Allan, "Cyclical Determinants of the Natural Level of Economic Activity," *International Economic Review*, June 1985, *26* (2), 387–97.

Dupraz, Stephane, Emi Nakamura, and Jon Steinsson, "A Plucking Model of Business Cycles," Technical Report 2017.

Eggertsson, Gauti B. and Michael Woodford, "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 2003, *34* (1), 139–235.

_ **and Neil R. Mehrotra**, "A Model of Secular Stagnation," NBER Working Papers 20574, National Bureau of Economic Research, Inc October 2014.

_ **and Paul Krugman**, "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach," *The Quarterly Journal of Economics*, 2012, *127* (3), 1469–1513.

Fallick, Bruce C., Daniel Villar Vallenas, and William L. Wascher, "Downward Nominal Wage Rigidity in the United States during and after the Great Recession," Working Papers 201602R, Federal Reserve Bank of Cleveland March 2020.

Galí, Jordi, "Insider-outsider labor markets, hysteresis and monetary policy," Technical Report January 2016.

Garga, Vaishali and Sanjay R. Singh, "Output Hysteresis and Optimal Monetary Policy," Technical Report November 2016.

Ghayad, Rand, "The Jobless Trap," Technical Report 2014.

Gordon, Robert J., "Secular Stagnation: A Supply-Side View," *American Economic Review*, May 2015, *105* (5), 54–59.

Grigsby, John, Erik Hurst, and Ahu Yildirmaz, "Aggregate Nominal Wage Adjustments: New Evidence from Administrative Payroll Data," NBER Working Papers 25628, National Bureau of Economic Research, Inc March 2019.

Guerrieri, Veronica and Guido Lorenzoni, "Credit Crises, Precautionary Savings, and the Liquidity Trap," *The Quarterly Journal of Economics*, 2017, *132* (3), 1427–1467.

Hall, Robert E, "Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor," *Journal of Political Economy*, 2009, *117* (2), 281–323.

Haltmaier, Jane, "Do recessions affect potential output?," International Finance Discussion Papers 1066, Board of Governors of the Federal Reserve System (U.S.) 2012.

Hazell, Jonathan and Bledi Taska, "Downward Rigidity in the Wage for New Hires," Technical Report 2019.

Kang, Dae Woong, Nick Ligthart, and Ashoka Mody, "The European Central Bank: Building a shelter in a storm," CFS Working Paper Series 527, Center for Financial Studies (CFS) 2015.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 2013, *128* (3), 1123–1167.

Laureys, Lien, "Optimal Monetary Policy in the Presence of Human Capital Depreciation during Unemployment," Discussion Papers 1415, Centre for Macroeconomics (CFM) June 2014.

Ljungqvist, Lars and Thomas J. Sargent, "The European Unemployment Dilemma," *Journal of Political Economy*, June 1998, *106* (3), 514–550.

Martin, Robert F., Tenyanna Munyan, and Beth Anne Wilson, "Potential Output and Recessions : Are We Fooling Ourselves?," IFDP Notes 2014-11-12, Board of Governors of the Federal Reserve System (U.S.) November 2014.

Menzio, Guido and Shouyong Shi, "Efficient Search on the Job and the Business Cycle," *Journal of Political Economy*, 2011, *119* (3), 468 – 510.

Murray, Seth, "Downward Nominal Wage Rigidity and Job Destruction," Technical Report 2019.

Paradise, Andrew, "Learning Remains Steady During the Downturn," https://www.td.org/Publications/Magazines/TD/TD-Archive/2009/11/Learning-Remains-Steady-During-the-Downturn November 2009.

Pissarides, Christopher A., "Loss of Skill During Unemployment and the Persistence of Employment Shocks," *The Quarterly Journal of Economics*, 1992, *107* (4), 1371–1391.

_ , "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?," *Econometrica*, 09 2009, *77* (5), 1339–1369.

Schaal, Edouard and Mathieu Taschereau-Dumouchel, "Coordinating Business Cycles," 2015 Meeting Papers 178, Society for Economic Dynamics 2015.

_ **and** _ , "Aggregate Demand and the Dynamics of Unemployment," Working Paper 2016.

Schmitt-Grohe, Stephanie and Martin Uribe, "Downward Nominal Wage Rigidity and the Case for Temporary Inflation in the Eurozone," *Journal of Economic Perspectives*, August 2013, *27* (3), 193–212.

_ **and** _ , "Liquidity Traps and Jobless Recoveries," *American Economic Journal: Macroeconomics*, January 2017, *9* (1), 165–204.

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

Song, Jae and Till von Wachter, "Long-Term Nonemployment and Job Displacement," Jackson Hole Economic Symposium Conference Proceedings, Kansas City Federal Reserve Board 2014.

Sterk, Vincent, "The Dark Corners of the Labor Market," Discussion Papers 1603, Centre for Macroeconomics (CFM) January 2016.

Wee, Shu Lin, "Delayed Learning and Human Capital Accumulation: The Cost of Entering the Job Market During a Recession," Technical Report 2016.

Werning, Ivan, "Managing a Liquidity Trap: Monetary and Fiscal Policy," NBER Working Papers 17344, National Bureau of Economic Research, Inc August 2011.

Yagan, Danny, "Employment Hysteresis from the Great Recession," Working Paper 23844, National Bureau of Economic Research September 2017.

Yellen, Janet, "Macroeconomic Research After the Crisis," https://www.federalreserve.gov/newsevents/speech/yellen20161014a.htm October 2016.

Appendix

A Wages and Nash bargaining

Recall that the value of an employed worker and of an unemployed worker, respectively, are defined by the recursions (2) and (3). Also, the value of a filled vacancy to a firm is given by equation (6). We can then define the surplus of a match between a worker and a firm as:

$$\mathcal{S}_t = J_t + \mathbb{W}_t - \mathbb{U}_t$$

Wages are determined by Nash bargaining. Denoting workers' bargaining power by η, wages solve

$$\max_{w_t} J_t^{1-\eta}(\mathbb{W}_t - \mathbb{U}_t)^\eta$$

implying

$$\eta J_t = (1-\eta)(\mathbb{W}_t - \mathbb{U}_t)$$

Notice that the match surplus can be rewritten as:

$$\begin{aligned}
\mathcal{S}_t &= J_t + \mathbb{W}_t - \mathbb{U}_t \\
&= A - b + \beta(1-\delta)J_{t+1} + \beta(1-\delta)(1-q_{t+1})(\mathbb{W}_{t+1} - \mathbb{U}_{t+1}) \\
&= A - b + \beta(1-\delta)(1-q_{t+1})S_{t+1} + \beta(1-\delta)q_{t+1}J_{t+1}
\end{aligned}$$

Using the fact that $\mathbb{W}_t - \mathbb{U}_t = \eta S_t$ in the equation above, we have:

$$\omega_t = \eta A + (1-\eta)b + \beta(1-\delta)\eta q_{t+1}J_{t+1}$$

B Existence of multiple steady states

Define:

$$\underline{\eta} = \max\left\{\frac{1-\beta(1-\delta)}{\beta(1-\delta)}, \frac{1-\delta}{2-\delta}, \frac{(1-\delta)(\kappa+\chi)-b}{(1-\delta)(\kappa+\chi)+a-b}\right\} \tag{20}$$

Steady states μ satisfies $\frac{a}{1-\beta(1-\delta)[1-\eta(1-\mu)]} = \kappa + \chi\mu$. Dividing through by J_{min}, this becomes

$$(1-e\mu)^{-1} = k + x\mu \tag{21}$$

where $k = \kappa/J_{min}$, $x = \chi/J_{min}$ and $e = \frac{\beta\eta(1-\delta)}{1-\beta(1-\delta)(1-\eta)}$. For future reference, we define $\underline{\chi} = e\left[2 - k + 2\sqrt{1-k}\right]$. Assumptions 1 and 2 impose that $k < 1$ and $(1-e)^{-1} > k + x$. Since $e \in (0,1)$, $(1-e\mu)^{-1}$ is an increasing, strictly convex function. Starting from $x = 0$, as we increase x, either the intersection of these two functions first occurs at $\mu \in (0,1)$, in which case a slightly higher x would give us multiplicity, or the first intersection has $\mu \geq 1$. Consider the knife edge case in which the first intersection of these two curves is at $\mu = 1$. Then the curves must be tangent and equal to each other at $\mu = 1$, i.e.

$$\frac{e}{(1-e)^2} = k \qquad \text{and} \qquad \frac{1}{1-e} = k + x$$

which implies $k = (1-2e)(1-e)^{-2}$.

In order to have multiple intersections in $(0,1)$, there must exist some $\mu \in (0,1)$ such that $(1-e\mu)^{-1} = k + x\mu$ and $e(1-e\mu)^{-2} > x$ (at the larger of the two intersections, this convex function must intersect the linear function from below). If $k < (1-2e)(1-e)^{-2}$, then this cannot be the case. A smaller k implies a larger x, increasing the slope of the linear function; $\mu < 1$ decreases the slope of the convex function. Thus, we must have $k > (1-2e)(1-e)^{-2}$.The assumption that $\eta \geq \underline{\eta}$ is sufficient (but not necessary) to ensure this, since it implies that $e > 0.5$. If this is true, and if x is just large enough that there is a single slack steady state, then (21), which is quadratic in μ, has a unique solution, i.e. its discriminant equals zero: $x^2 - 2e(2-k)x + e^2k^2 = 0$.

Considered as a function of x, this equation has two real solutions since its discriminant is positive: $4(e^2(2-k)^2 - e^2k^2) = 16e^2(1-k) > 0$. This will have two solutions x^*, the larger of which corresponds to $\mu \in (0,1)$. To see this, consider the following graphical argument. Fix e and $k < 1$ and start with $x = \infty$, so that the $k + x\mu$ line is vertical at $\mu = 0$. Then the two curves intersect at exactly one point, $\mu = 0$. Decreasing x rotates the straight line clockwise, increasing the smallest value of μ at which the two curves intersect from 0 to some positive number. Eventually, for low enough x, the straight line is tangent to the convex curve at some $\mu > 0$. Next, start with $x = 0$, so the straight line $k + x\mu$ is horizontal at k and intersects the convex curve at some $\mu = e^{-1}(1 - k^{-1}) < 0$. Gradually increasing x rotates the straight line counter-clockwise, lowering the first value at which the curves intersect. For x large enough, the two curves are tangent at some $\mu < 0$. Clearly, the second case corresponds to a lower value of x. Thus, the larger value of x corresponds to the economically sensible case where $\mu \in (0,1)$. Choosing this value, we have

$$x^* = e(2-k) + \sqrt{e^2(2-k)^2 - e^2k^2} = e[2 - k + 2\sqrt{1-k}]$$

Thus there will be multiple steady states if $x > x^*$.

C Proof of Proposition 1

Suppose $\mu_0 = 0$. Then, note that $\mu_t = 0$ (which implies $n_t = 1$) is consistent with (10), since in the tight labor market regime $q_t = 1$, and $n_t = 1, \mu_{t+1} = 0$. Next we show that we cannot have $\theta_0 < \theta^{\text{fe}}$ given $\mu_0 = 0$. (Since $\theta^{\text{fe}} \geq 1$ by Assumption 1, this implies in particular that we cannot have $\theta_0 < 1$.) In any equilibrium, (7) must be satisfied:

$$J_{min} \leq a \sum_{t=0}^{\infty} \prod_{\tau=0}^{t} \beta(1-\delta)(1 - \eta \min\{\theta_\tau, 1\}) \leq \kappa \max\{\theta_t, 1\}$$

where the first inequality holds because the LHS is decreasing in θ_τ. Since we know that $J_{min} > \kappa$ from Assumption 1, it is immediate that this inequality can only be satisfied if $\theta_t \geq \theta^{\text{fe}} \geq 1$. Finally, we show that we cannot have $\theta_0 > \theta^{\text{fe}}$. We have shown that

$$a \sum_{t=0}^{\infty} \prod_{\tau=0}^{t} \beta(1-\delta)(1 - \eta \min\{\theta_\tau, 1\}) = \kappa\theta_0$$

in any equilibrium, and that this expression is satisfied by $\theta_t = \theta^{\text{fe}}$, $\forall t \geq 0$. If $\theta_0 > \theta^{\text{fe}}$, it follows that $\theta_t < \theta^{\text{fe}}$ for some $t > 0$. Let T be the first date at which this is true. Then up to that date, since the labor market has been tight, $\mu_T = 0$. This is a contradiction, since we have already shown that if $\mu_T = 0$, $\theta_T \geq \theta^{\text{fe}}$. It follows that the unique equilibrium has $\theta_t = \theta^{\text{fe}}$ for all $t \geq 0$. The proof for any $\mu_0 \in (\mu_0, \underline{\mu})$ is similar and follows from the fact that $q_0 = 1$ which implies that all workers are employed in period 0. Before characterizing the case when $\mu = \underline{\mu}$, the following result is useful:

Lemma 1. *If $J_t = J_{min}$, then $q_{t+1} = 1$, i.e. $\theta_{t+1} \geq 1$ and $J_{t+1} = J_{min}$.*

Proof. We have $J_t = a + \beta(1-\delta)(1-\eta q_{t+1})J_{t+1}$. The only way to attain $J_t = J_{min}$ is $q_{t+1} = 1$ and $J_{t+1} = J_{min}$, since $q_{t+1} \leq 1$, $J_{t+1} \geq J_{min}$, and the expression is decreasing in q_{t+1} and increasing in J_{t+1}. $\square$

For $\mu_0 = \underline{\mu}$, there exist a continuum of equilibria indexed by $\theta_0 \in [1-\underline{\mu}, 1]$. In all these equilibria, the value of an employed worker for a firm is given by J_{min}. To see this, notice that $J_0 \leq \kappa + \chi\underline{\mu}$ as long as labor markets are slack, $\theta_0 \leq 1$. In this case, by definition, $J_0 \leq J_{min}$ and by definition this relationship has to hold with equality. If labor markets are tight, $\theta_0 > 1$, then $J_0 = \kappa\theta_0 + \chi\underline{\mu} > J_{\min}$ since $\theta_0 > 1$. This is a contradiction since if $\theta_0 > 1$, $\mu_1 = 0$ from Lemma 1 and $J_0 = J_{min}$ from Lemma 1. Furthermore, from Lemma 1, it follows that $J_1 = J_{min}$ and $\theta_1 \geq 1$.

The contradiction above shows that $\theta_0 \leq 1$. We now need to show that $\theta_0 > 1 - \underline{\mu}$. Suppose that $\theta_0 < 1 - \underline{\mu}$. Then μ_1 is given by:

$$\mu_1 = \frac{1-\theta_0}{1+(1-\delta)[1-\theta_0-\underline{\mu}]} > \underline{\mu}$$

This is a contradiction since

$$J_1 = J_{min} = \kappa + \chi\underline{\mu} < \kappa + \chi\mu_1$$

which requires that $\theta_1 = 0$. Thus, we have shown that $\theta_0 \in [1-\underline{\mu}, 1]$. From (4) and the earlier part of this proof, it follows that $\mu_1 = \frac{1-\theta_0}{1+(1-\delta)[1-\theta_0-\mu_0]} \leq \mu_0$ and $\theta_1 = (J_{min} - \chi\mu_1)/\kappa \geq 1$. As mentioned in footnote 8, we select the equilibrium in which $\theta_0 = 1$ implying that $\mu_1 = 0$, $\theta_1 = \theta^{\text{fe}}$.

D Proof of Proposition 2

Definition 1. *Define the functions $\Theta^1 : I^1 \to [0,1]$, $F^1 : I^1 \to \mathbb{R}_+$, $M^1 : I^1 \to \{\underline{\mu}\}$ as:*

$$\begin{aligned}
\Theta^1(\mu_{T-1}) &:= 1 - \frac{\underline{\mu}}{1-(1-\delta)\underline{\mu}}(1-(1-\delta)\mu_{T-1}) \\
F^1(\mu_{T-1}) &:= \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta\Theta^1(\mu_{T-1}))(\kappa + \chi\mu_{T-1})\right] \\
M^1(\mu_{T-1}) &:= \underline{\mu}
\end{aligned}$$

where $I^1 = [\underline{\mu}, \mu^1]$ and $\mu^1 := F^1(\underline{\mu})$.

Intuitively, at any date t, for any $\mu_t \in I^1$, $\Theta^1(\mu_t)$ describes the job-finding rate that ensures that the economy reaches $\underline{\mu}$ at date $t+1$. $F^1(\mu_t)$ describes the unique value that μ_{t-1} can have in period $t-1$ such that $\mu_t \in I^1$ and also $\mu_{t+1} = \underline{\mu}$. In other words, given market tightness at date t, $\Theta^1(\mu_t)$, one can compute the value of a filled vacancy at date $t-1$ and zero and by no-arbitrage, this pins down the value of μ_{t-1} for which firms would have been willing to post the requisite number of vacancies. $M^1(\mu)$ is just a constant function which by definition describes where any $\mu \in I^1$ ends up.

Corollary 1. *It must be true that* $\mu^1 < \widetilde{\mu}$.

By the definition of μ^1, it must be true that

$$\begin{aligned} \mu^1 &= \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)\left[1-\eta(1-\underline{\mu})\right](\kappa + \chi\underline{\mu})\right] \\ &< \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)\left[1-\eta(1-\widetilde{\mu})\right](\kappa + \chi\widetilde{\mu})\right] \\ &= \widetilde{\mu} \end{aligned}$$

Lemma 2. *For* β *sufficiently close to 1,* F^1 *is increasing in* μ *for* $\mu \in [\underline{\mu}, \widetilde{\mu})$

Proof. Since $F^1(\mu)$ is composed of constants and a concave part, it suffices to consider the concave polynomial $\xi(\mu) = \left[1 - \eta\Theta^1(\mu)\right](\kappa + \chi\mu)$. This function is increasing in μ for

$$\mu < \frac{1}{2}\left[\frac{\left(1-(1-\delta)\underline{\mu}\right)(1-\eta)}{\eta(1-\delta)\underline{\mu}} + \frac{1}{(1-\delta)} - \frac{\kappa}{\chi}\right] \tag{22}$$

It is thus sufficient to show that $\widetilde{\mu}$ satisfies this inequality. Before proceeding further, it is convenient to work with a quasi-value function of the firm defined in terms of μ as opposed to J_t. Define the quasi-value function $\mathcal{Q}(\mu)$ as:

$$\mathcal{Q}(\mu) = \frac{a}{1-\beta(1-\delta)\left[1-\eta(1-\mu)\right]}$$

By construction, $\mathcal{Q}(\mu)$ is the value of the firm as long as the job-finding rate is $1-\mu$ forever. Note that $\mathcal{Q}'(\mu) > 0$ and $\mathcal{Q}''(\mu) > 0$.

Under this quasi-value function and given free entry, $\widetilde{\mu}$ satisfies

$$\mathcal{Q}(\widetilde{\mu}) = \frac{a}{1-\beta(1-\delta)\left(1-\eta+\eta\widetilde{\mu}\right)} = \kappa + \chi\widetilde{\mu}$$

Since the left hand side is convex and the right hand side linear, since $\widetilde{\mu}$ is the smaller of two solutions to this equation, then

$$\mathcal{Q}'(\widetilde{\mu}) = \frac{a\beta(1-\delta)\eta}{\left[1-\beta(1-\delta)\left(1-\eta+\eta\widetilde{\mu}\right)\right]^2} < \chi$$

In other words, the LHS cuts the RHS from above. Next, dividing the first equality by the second inequality, we have

$$\widetilde{\mu} < \frac{1}{2}\left[\frac{1-\beta(1-\delta)}{\beta(1-\delta)\eta} + 1 - \frac{\kappa}{\chi}\right] \tag{23}$$

Define:

$$\Xi = \frac{1}{2}\left\{\frac{(1-(1-\delta)\underline{\mu})(1-\eta)}{\eta(1-\delta)\underline{\mu}} - \frac{1-\beta(1-\delta)}{\eta\beta(1-\delta)} + \frac{1}{(1-\delta)} - 1\right\}$$

Assuming that $\beta > \frac{\widetilde{\mu}}{\eta\widetilde{\mu}+1-\eta}$, it can be shown that $\Xi > 0$.[23] Thus, as required:

$$\widetilde{\mu} < \frac{1}{2}\left[\frac{1-\beta(1-\delta)}{\beta(1-\delta)\eta} + 1 - \frac{\kappa}{\chi}\right] + \Xi = \frac{1}{2}\left[\frac{\left(1-(1-\delta)\underline{\mu}\right)(1-\eta)}{\eta(1-\delta)\underline{\mu}} + \frac{1}{(1-\delta)} - \frac{\kappa}{\chi}\right]$$

$\square$

It was already clear that given a $\mu_{t+1} \in I^1$, there exists a unique μ_t which could have led there. In addition, this Lemma shows that given any μ_t, there exists at most one $\mu_{t+1} \in I^1$ is consistent with equilibrium.

Corollary 2. *Let $I^2 = F^1(I^1)$ and let $M^2(\mu)$ be the inverse of this function. Then $M^2(\mu^1) = M^1(\mu^1) = \underline{\mu}$.*

Since F^1 is increasing and continuous, its inverse M^2 exists and is increasing and continuous. Consequently, $F^1(I^1)$ maps into an interval $(\mu^1, \mu^2]$. Further since $\mu^1 = F^1(\underline{\mu})$, then $M^2(\mu^1) = \underline{\mu}$.

Lemma 3. $\mu^2 = F^1(\mu^1) < \widetilde{\mu}$

Proof. Since $\Theta^1(\underline{\mu}) = 1 - \underline{\mu}$ and Θ^1 is increasing, we have $\Theta^1(\mu^1) > 1 - \underline{\mu} > 1 - \widetilde{\mu}$. It follows that:

$$\frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta\Theta^1(\mu^1))(\kappa + \chi\widetilde{\mu})\right] \quad < \quad \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta(1-\widetilde{\mu}))(\kappa + \chi\widetilde{\mu})\right]$$

Then, from Corollary 1, since $\mu^1 < \widetilde{\mu}$:

$$\begin{aligned} F^1(\mu^1) &= \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta\Theta^1(\mu^1))(\kappa + \chi\mu^1)\right] \\ &< \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta(1-\widetilde{\mu}))(\kappa + \chi\widetilde{\mu})\right] \\ &= \widetilde{\mu} \end{aligned}$$

$\square$

Lemma 4. *Define $\Theta^2(\mu) : I^2 \to [0,1]$ as:*

$$\Theta^2(\mu) := 1 - M^2(\mu)\frac{1-(1-\delta)\mu}{1-(1-\delta)M^2(\mu)}$$

Then,

$$\frac{\partial\Theta^2(\mu)}{\partial\mu} \leq \frac{(1-\delta)M^2(\mu)}{1-(1-\delta)M^2(\mu)}$$

[23] Note that this assumption is a condition on an endogenous variable, $\widetilde{\mu}$ and can be rewritten as $\widetilde{\mu} < \dfrac{1-\eta}{\beta^{-1}-\eta}$. Nonetheless, it is a weak condition: for any $\widetilde{\mu} < 1$, it is satisfied for β sufficiently close to 1.

Proof.

$$\begin{aligned}\frac{\partial\Theta^2(\mu)}{\partial\mu} &= M^2(\mu)\frac{(1-\delta)}{1-(1-\delta)M^2(\mu)} - \frac{\partial M^2(\mu)}{\partial\mu}\left[1+\frac{(1-\delta)\left(1-(1-\delta)\mu\right)M^2(\mu)}{\left[1-(1-\delta)M^2(\mu)\right]^2}\right] \\ &\leq M^2(\mu)\frac{(1-\delta)}{1-(1-\delta)M^2(\mu)}\end{aligned}$$

where the inequality comes because $M^2(\mu)$ is increasing and the expression in square brackets is positive. □

We are now ready to characterize equilibrium in the entire convalescent region.

Lemma 5 (Induction Step)**.** *Suppose the functions $\Theta^n(\mu)$, $M^n(\mu)$ are defined on some interval $I^n = [\mu^{n-1}, \mu^n]$ and $M^{n-1}(\mu_{T-n+1})$ is defined on an interval $I^{n-1} = [\mu^{n-2}, \mu^{n-1}]$, with $\underline{\mu} < \mu^{n-2} < \mu^n < \widetilde{\mu}$, and that these functions satisfy*

$$\begin{aligned}\Theta^n(\mu) &= 1 - M^n(\mu)\frac{1-(1-\delta)\mu}{1-(1-\delta)M(\mu)} \\ \frac{\partial\Theta^n(\mu)}{\partial\mu} &< \frac{(1-\delta)M^n(\mu)}{1-(1-\delta)M^n(\mu)} \\ M^n(I^n) &= I^{n-1} \\ M^n(\mu^{n-1}) &= M^{n-1}(\mu^{n-1}) = \mu^{n-2}\end{aligned}$$

Then, for β sufficiently close to 1, we have the following results:

1. *The function*

$$F^n(\mu) := \frac{1}{\chi}\left[a - \kappa + \beta(1-\delta)(1-\eta\Theta^n(\mu))(\kappa+\chi\mu)\right]$$

   is monotonically increasing in μ for $\mu \leq \widetilde{\mu}$.

2. *Let $I^{n+1} = F^n(I^n)$ and let $M^{n+1}(\mu)$ be the inverse of this function. Then $M^{n+1}(\mu^n) = M^n(\mu^n) = \mu^{n-1}$.*

3. *$I^{n+1} = [\mu^n, \mu^{n+1}]$ with $\mu^{n+1} < \widetilde{\mu}$.*

4. *Define $\Theta^{n+1}(\mu)$ on I^{n+1} by*

$$\Theta^{n+1}(\mu) = 1 - M^{n+1}(\mu)\frac{1-(1-\delta)\mu}{1-(1-\delta)M^{n+1}(\mu)}$$

   The derivative of this function satisfies

$$\frac{\partial\Theta^{n+1}(\mu)}{\partial\mu} < \frac{(1-\delta)M^{n+1}(\mu)}{1-(1-\delta)M^{n+1}(\mu)}$$

Proof. (1.) The derivative of $F^n(\mu)$ is

$$\begin{aligned}\frac{\partial F^n(\mu)}{\partial \mu} &= \frac{\beta(1-\delta)}{\chi}\left[-\eta\frac{\partial \Theta^n(\mu)}{\partial \mu}(\kappa+\chi\mu)+\chi(1-\eta\Theta^n(\mu))\right] \\ &> \frac{\beta(1-\delta)}{\chi}\left[-\eta\frac{(1-\delta)M^n(\mu)}{1-(1-\delta)M^n(\mu)}(\kappa+\chi\mu)+\chi(1-\eta\Theta^n(\mu))\right]\end{aligned}$$

Substituting in the definition of Θ^n and rearranging, we see that this expression will be positive provided that

$$\mu < \frac{1}{2}\left[\frac{1-\eta}{\eta}\frac{(1-(1-\delta)M^n(\mu))}{(1-\delta)M^n(\mu)}+\frac{1}{(1-\delta)}-\frac{\kappa}{\chi}\right]$$

By the same logic as in Lemma 2, for β sufficiently close to 1, this is satisfied for any $\mu \leq \widetilde{\mu}$, since we have $M^n(\mu) \leq \widetilde{\mu}$. So $F^n(\mu)$ is increasing, and hence invertible, for $\mu < \widetilde{\mu}$. Let $M^{n+1}(\mu)$ be the inverse of this function.

(2.) We have

$$\begin{aligned} M^n(\mu^{n-1}) &= M^{n-1}(\mu^{n-1}) \\ \Theta^n(\mu^{n-1}) &= \Theta^{n-1}(\mu^{n-1}) \\ F^n(\mu^{n-1}) &= F^{n-1}(\mu^{n-1}) = \mu^n \text{ by definition of } \mu^n \\ M^{n+1}(\mu^n) &= M^n(\mu^n)\end{aligned}$$

(3.) Since F^n is a continuous, increasing function, the image of the interval $[\mu^{n-1}, \mu^n]$ under F^n must be an interval $[\mu^n, \mu^{n+1}]$. (We have already shown that $F^n(\mu^{n-1}) = \mu^n$.) We need to show that $\mu^{n+1} = F^n(\mu^n) < \widetilde{\mu}$. We know that $\widetilde{\mu} \geq M^n(\mu^n)$. Then, it must be true that

$$\begin{aligned} 1-\widetilde{\mu} &< 1-M^n(\mu^n) \\ &= 1-M^n(\mu^n)\frac{1-(1-\delta)\mu^n}{1-(1-\delta)\mu^n} \\ &< 1-M^n(\mu^n)\frac{1-(1-\delta)\mu^n}{1-(1-\delta)M^n(\mu^n)} \\ &= \Theta^n(\mu^n)\end{aligned}$$

Then, by the same logic as in Lemma 3 we have $F^n(\mu^n) < \widetilde{\mu}$. So we have shown that $I^{n+1} \subset [\underline{\mu}, \widetilde{\mu}]$.

(4.) The bound on the derivative is established in the same way as Lemma 4.

□

Lemma 6. $\lim_{n\to\infty} \mu^n \to \widetilde{\mu}$.

Proof. We have shown that $\{\mu^n\}$ is an increasing sequence bounded above by $\widetilde{\mu}$; thus by the Monotone Convergence Theorem, its limit μ^∞ exists, and $\mu^\infty \leq \widetilde{\mu}$. Suppose by contradiction that $\mu^\infty < \widetilde{\mu}$. Then μ^∞ must be a steady state. But by definition, $\widetilde{\mu}$ is the smallest slack steady state. So we must have $\mu^\infty = \widetilde{\mu}$. □

Finally, we prove that recoveries can be arbitrarily slow, i.e. for any $T \in \mathbb{N}$, there exists $\varepsilon > 0$ such that if $\mu_0 \in (\widetilde{\mu} - \varepsilon, \widetilde{\mu})$, $\mu_t > 0$ for all $t < T$. Fix $\delta > 0, T \in \mathbb{N}$ and let n be the smallest integer such that $\mu^n \geq \widetilde{\mu} - \delta$ (this exists, since $\mu^n \to \widetilde{\mu}$ and $\delta > 0$. Set $\varepsilon = \widetilde{\mu} - \mu^{n+T}$. Take any $\mu_0 \in (\widetilde{\mu} - \varepsilon, \widetilde{\mu}) = (\mu^{n+T}, \widetilde{\mu})$. Then $\mu_0 \in (\mu^{m-1}, \mu^m]$ for some $m > n + T + 1$. We know from 2 that $\mu_T \in (\mu^{m-T-1}, \mu^{m-T}]$. In particular,

$$\mu_T > \mu^{m-T-1} > \mu^n \geq \widetilde{\mu} - \delta > \mu_0 - \delta$$

Finally, since $\{\mu_t\}$ is monotonically decreasing, we have $\mu_t > \mu_0 - \delta$ for all $t < T$, as claimed. Next, note that the first part of the lemma is a special case of the second part with $\delta = \widetilde{\mu}$.

E Proof of Proposition 3

To see this more formally, note that any trajectory which starts to the right of $\underline{\mu}$ and reached full employment at some date T has to be at $\underline{\mu}$ at date $T-2$. But Proposition 2 showed that *all* trajectories that reach $\underline{\mu}$ lie entirely within the convalescent region. It follows that if the economy starts in the stagnant region - defined as the set $[\widetilde{\mu}, 1]$ - it can never converge to full employment - this region is an *unemployment trap*.

F Planning Problems

Discretion Problem

We consider a planner who solves the problem

$$\mathbb{L}\left(\mu_t, W_{t-1} \mid \beta_t\right) = \min_{W_t, u_t, P_t, \theta, \mu_{t+1}} u_t^2 + \lambda \left(\frac{W_t}{W_{t-1}} - 1\right)^2 + \beta_t \mathbb{L}\left(\mu_{t+1}, W_t \mid \beta_{t+1}\right)$$

s.t.

$$\begin{aligned}
W_t &= \max\left\{\varphi W_{t-1}, P_t \omega^*\left(\mu_t\right)\right\} \\
\mu_{t+1} &= \frac{u_t}{1 - (1-\delta)\left(1 - u_t\right)} \\
\frac{\kappa}{f_t} + \chi \mu_t &= A - \frac{W_t}{P_t} + \beta_t (1-\delta)\left(\frac{\kappa}{f_{t+1}} + \chi \mu_{t+1}\right) \\
\mu_{t+1} &= \frac{1 - q_t}{1 + (1-\delta)\left(1 - q_t - \mu_t\right)} \\
f_t &= \min\left\{\frac{1}{\theta_t}, 1\right\}, \qquad q_t = \min\left\{1, \theta_t\right\}, \qquad P_t \leq \frac{1}{\beta_t} P_{t+1}
\end{aligned}$$

where $\beta_0 > 1$, $\beta_t = \beta \in (0,1)$ for all $t > 0$. We will show that the solution to this problem is

$$P_t \quad \leq \quad \frac{W_{t-1}}{\omega^*(\mu_t)},\ i_t \geq 0 \qquad \text{with at least one equality}$$

and if $P_t = \frac{W_{t-1}}{\omega^*(\mu_t)}$, then $\mu_{t+1} = \mathcal{M}(\mu_t)$.

Suppose that monetary policy indeed follows the policy described above from date $t+1$ onwards. This implies that the loss $\mathbb{L}(\mu_{t+1}, \cdot \mid \beta_{t+1})$ is increasing in μ_{t+1}, since a higher value of μ_{t+1} implies a higher value of μ_{t+k} for all $k > 1$, and thus a larger per-period loss u^2. Thus at date t, choosing $\mu_{t+1} > \mathcal{M}(\mu_t)$ delivers a strictly higher loss than choosing $\mu_{t+1} = \mathcal{M}(\mu_t)$ and $W_t = W_{t-1}$. If setting $P_t = \frac{W_{t-1}}{\omega^*(\mu_t)}$ and $\mu_{t+1} = \mathcal{M}(\mu_t)$ does not violate the ZLB, it is therefore optimal. Setting a price $P_t < \frac{W_{t-1}}{\omega^*(\mu_t)}$ is not optimal as it results in a higher real wage, fewer vacancies and thus a higher μ_{t+1} which entails a higher loss, as we have just argued. Finally, because real wages cannot be lower than their natural level, so setting a higher price $P_t > \frac{W_{t-1}}{\omega^*(\mu_t)}$ not only still leads to $\mu_{t+1} = \mathcal{M}(\mu_t)$ but also gives rise to a non-zero wage inflation $W_t > W_{t-1}$ as nominal wages increase one-for-one with prices in this region. It follows that any positive level of wage inflation leads to a strictly higher loss. Thus, implementing zero nominal wage inflation and replicating the natural allocation is optimal under discretion as long as the ZLB does not bind.

However, if setting $P_t = \frac{W_{t-1}}{\omega^*(\mu_t)}$ and $\mu_{t+1} = \mathcal{M}(\mu_t)$ does violate the ZLB at date t, then optimal policy under discretion will implement $i_t = 0$ and $P_t < \frac{W_{t-1}}{\omega^*(\mu_t)}$. It can never be optimal to set $i_t = 0$ and $P_t > \frac{W_{t-1}}{\omega^*(\mu_t)}$: whenever it is feasible to do so, implementing the natural real wage and zero wage inflation is also feasible and optimal. Setting $i_t = 0$ and $P_t > \frac{W_{t-1}}{\omega^*(\mu_t)}$ would result in positive wage inflation $W_t > W_{t-1}$ and real wages at their flexible price level $\frac{W_t}{P_t} = \omega^*(\mu_t)$. This is dominated by setting $P_t = \frac{W_{t-1}}{\omega^*(\mu_t)}$, which results in the same level of real wage and thus the same μ_{t+1}, but zero nominal wage inflation. Implementing this lower price level does not violate the ZLB as it entails setting a higher interest rate $i_t > 0$. Thus, whenever it is constrained optimal to set $i_t = 0$, it must be that $P_t < \frac{W_{t-1}}{\omega^*(\mu_t)}$. □

Commitment

We now show that UMP is optimal under commitment for a planner who seeks to minimize

$$\sum_{t=0}^{\infty} \prod_{j=0}^{t} \beta_{t-j} \left\{ u_t^2 + \lambda \left(\frac{W_t}{W_{t-1}} - 1 \right)^2 \right\}$$

for $\lambda > 0$ sufficiently small. Under UMP, $\frac{W_1}{W_0} = \beta_0 \frac{\omega^*_{\text{fe}}}{\omega_0^{ump}} > 1$ at date 1; wage inflation equals zero at all other dates, and unemployment remains at zero at all dates. Thus, UMP attains the loss $\beta_0 \lambda \left(\beta_0 \frac{\omega^*_{\text{fe}}}{\omega_0^{ump}} - 1 \right)^2$. Since unemployment remains at zero under UMP, to show that this loss is smaller than the loss associated with any other policy, it suffices to compare it to other policies which involve less nominal wage inflation at date 1, i.e.,

$$\beta_0 \frac{\omega^*_{\text{fe}}}{\omega_0^{ump}} > \frac{W_1}{W_0} = \frac{P_1}{P_0}\frac{\omega_1}{\omega_0} \geq \beta_0 \frac{\omega_1}{\omega_0}.$$

That is, under one of these alternative policies, either $\omega_0 > \omega_0^{ump}$, or $\omega_1 < \omega^*_{\text{fe}}$, or both. Either one of these conditions implies that we must have $\mu_1 \geq \underline{\mu}$. First suppose $\omega_1 < \omega^*_{\text{fe}}$. Since $\omega_1 \geq \omega^*(\mu_1)$ and $\omega^*(\mu) = \omega^*_{\text{fe}}$ for all $\mu < \underline{\mu}$, this directly implies $\mu_1 \geq \underline{\mu}$. Next, suppose $\omega_0 > \omega_0^{ump}$. The free entry condition at date 0 is

$$J_0 = A - \omega_0 + \beta_0(1-\delta)J_1 \quad \geq \quad \kappa = A - \omega_0^{ump} + \beta_0(1-\delta)J_{min},$$

which can be rearranged to get:

$$J_1 - J_{min} \geq \frac{\omega_0 - \omega_0^{ump}}{\beta_0(1-\delta)} > 0.$$

So we have $J_1 > J_{min}$. Given the definitions of these variables,

$$J_1 = \sum_{t=1}^{\infty} \beta^{t-1}(1-\delta)^{t-1}(A-\omega_t), \quad \text{and} \quad J_{min} = \frac{A - \omega^*_{\text{fe}}}{1-\beta(1-\delta)}$$

this implies that there must be some date $T \geq 1$ at which $\mu_T \geq \underline{\mu}$ for the first time. (If this were not true, and $\mu_t < \underline{\mu}$ for all t, then since $\omega_t \geq \omega^*(\mu_t)$ and $\omega^*(\mu) = \omega^*_{\text{fe}}$ for all $\mu < \underline{\mu}$, we would have $J_1 = J_{min}$.) We will now show that we must have $T = 1$. Suppose by contradiction that $T > 1$; then $\mu_T > 0$ implies $\theta_{T-1} < 1$ and so the date $T-1$ free entry condition is $J_{T-1} \leq \kappa < J_{min}$. But then, since $\mu_t < \underline{\mu}$ and $\omega_t \geq \omega^*_{\text{fe}}$ for all $1 \leq t < T$, we have

$$J_1 = \sum_{t=1}^{T-2} [\beta(1-\delta)]^{t-1}(A-\omega_t) + [\beta(1-\delta)]^{T-1} J_{T-1} < J_{min},$$

which contradicts the condition that $J_1 > J_{min}$. So we must have $\mu_1 \geq \underline{\mu}$ under any policy involving less date 1 wage inflation than UMP, as claimed above. Thus, date 0 unemployment must be at least $u(\underline{\mu}) \equiv \frac{\delta\underline{\mu}}{1-\underline{\mu}(1-\delta)}$, and the loss from this alternative poicy must be at least $u(\underline{\mu})^2$. Thus, UMP will be preferred to this alternative policy as long as $\beta_0\lambda\left(\beta_0\frac{\omega^*_{\text{fe}}}{\omega_0^{ump}} - 1\right)^2 \leq u(\underline{\mu})^2$ which is the case for λ small enough:

$$0 \leq \lambda \leq \frac{u\left(\underline{\mu}\right)^2}{\left(\beta_0\frac{\omega^*_{\text{fe}}}{\omega^*_0} - 1\right)^2}.$$

G Properties of $J_0(\beta_0)$

Suppose that the economy remains at full employment steady state even after the shock $\beta_0 > 1$. There are two cases to consider. First, suppose that the ZLB does not bind at date 0. Then monetary policy is unconstrained in all periods, and nominal wages and prices remain constant. From (14), we have $1 + i_t = \frac{P_1}{P_0 \beta_0} = \frac{1}{\beta_0}$. When $\beta_0 > 1$, this would imply a negative nominal interest rate, violating the ZLB. Thus, when $\beta_0 > 1$, monetary policy is constrained at date 0 and we have $P_0 = \frac{P_1}{\beta_0}$. Since the economy returns to full employment after date 0, real wages will equal ω^*_{fe} at all dates $t \geq 1$. Iterating forward $P_t = \min\left\{\frac{W_{t-1}}{\omega^*(\mu_t)}, \frac{P_{t+1}}{\beta_t}\right\}$, it follows that prices and nominal wages remain constant thereafter and the ZLB does not bind after date 0. In particular, since $W_1 = W_0$, we have:

$$\omega_0 = \frac{W_0}{W_1}\frac{P_1}{P_0}\omega_1 = \beta_0 \omega^*_{\text{fe}}$$

Using this in the expression for J_0 we have:

$$J_0 = A - \beta_0 \omega^*_{\text{fe}} + \beta_0(1-\delta)J_{min}$$

The full employment steady state Nash wage equals

$$\omega^*_{\text{fe}} = \frac{\eta}{1-\beta(1-\delta)(1-\eta)}A + \frac{[1-\beta(1-\delta)](1-\eta)}{1-\beta(1-\delta)(1-\eta)}b$$

So

$$\frac{\partial J}{\partial \beta_0} = -\omega^*_{\text{fe}} + (1-\delta)J_{min} = -\frac{\eta}{1-\beta(1-\delta)(1-\eta)}A - \frac{[1-\beta(1-\delta)](1-\eta)}{1-\beta(1-\delta)(1-\eta)}b + (1-\delta)\frac{(1-\eta)(A-b)}{1-\beta(1-\delta)(1-\eta)}$$

which is negative provided that $A\left[1-\delta-\frac{\eta}{1-\eta}\right] - [2-\delta-\beta(1-\delta)]b < 0$. By Assumption 2, both terms are negative, so this condition is satisfied.

H Proof of Proposition 4

First we show that a one-period hiring freeze takes the economy either to the convalescent or to the stagnant region.

Lemma 7. *Starting from full employment, a one period hiring freeze takes the economy out of the healthy region:* $\mu_R = \dfrac{1}{2-\delta} > \underline{\mu}$.

Proof. We prove the Lemma by proving the contrapositive. The first thing to note is that $\mu_R := \frac{1}{2-\delta} > 0.5$ since $0 < 1-\delta < 1$. Recall that $\underline{\mu} = \frac{J_{min}-\kappa}{\chi}$. Suppose $\underline{\mu} \geq \mu_R$. This implies that $\underline{\mu}$ must also be greater than 0.5. In this case, no interior steady state can exist. Recall that any interior steady state solves:

$$\begin{aligned}
\kappa + \chi\mu &= \mathcal{Q}(\mu) \\
&= \frac{a}{1-\beta(1-\delta)[1-\eta(1-\mu)]} \\
&= \frac{a}{1-\beta(1-\delta)(1-\eta)} \frac{1-\beta(1-\delta)(1-\eta)}{1-\beta(1-\delta)[1-\eta(1-\mu)]} \\
&= J_{min} \frac{1}{1-e\mu}
\end{aligned}$$

where, as before $e = \frac{\beta(1-\delta)\eta}{1-\beta(1-\delta)(1-\eta)}$.

Thus interior steady states solve:

$$\Omega(\mu) := \frac{J_{min}}{1-e\mu} - \kappa - \chi\mu = 0$$

We show that this is not possible if $\underline{\mu} > \mu_R$. In particular, we have $\Omega(\mu) > 0$ for all $\mu \in [0, 1]$. First, we show that $e > 1/2$ and $\chi < 2(J_{min} - \kappa)$. Notice that e can also be rewritten as:

$$e = \frac{1}{1 + \frac{1-\beta(1-\delta)}{\beta(1-\delta)\eta}} > \frac{1}{1 + \frac{\eta}{\eta}} = \frac{1}{2}$$

where the inequality follows since $\eta > \frac{1-\beta(1-\delta))}{\beta(1-\delta)}$ by Assumption 2. Thus, $e > \frac{1}{2}$. To see that $\chi < 2(J_{min} - \kappa)$, note that from the definition of $\underline{\mu}$:

$$\chi = \frac{J_{min} - \kappa}{\underline{\mu}} < 2(J_{min} - \kappa)$$

since $\underline{\mu} > 0.5$ by assumption.

Fix $\kappa \in [0, J_{min})$, $\mu \in [0, 1]$. Even though we have shown above that $e > 1/2$ and $\chi < 2(J_{min} - \kappa)$, for a moment, set $e = 1/2$, $\chi = 2(J_{min} - \kappa)$. We claim that

$$\mathcal{Q}(\mu) = \frac{J_{min}}{1-e\mu} \geq \kappa + \chi\mu = \kappa + 2(J_{min} - \kappa)\mu$$

with strict inequality unless $\kappa = 0$ and $\mu = 1$, in which case the expression holds with equality. When $\kappa = 0$, the RHS becomes $2J_{min}\mu$, and the LHS and RHS are only equal for $\mu = 1$. For any $\mu < 1$, the LHS is larger. When $\kappa > 0$, the RHS is strictly lower for any $\mu > 1/2$. Thus for any $\kappa \in [0, J_{min}]$, the inequality holds for all $\mu \in [0, 1)$. Finally, for any $\mu \leq 1/2$, the inequality clearly holds since the LHS is greater than J_{min}, and the RHS smaller than J_{min}.

Next, suppose $e > 1/2$ and $\chi < 2(J_{min} - \kappa)$. If $\mu = 0$, this does not change the inequality, which still holds strictly (since $\mu \neq 1$). If $\mu > 0$, this strictly increases the LHS and strictly decreases the RHS. Thus the expression is still satisfied with strict inequality. Thus we have $\Omega(\mu) > 0$ for all $\mu \in [0, 1]$, and there is no interior steady state. Since we have shown that $\underline{\mu} \geq \mu_R$ implies there exists no interior steady state, it follows that if there exist multiple interior steady states, we must have $\underline{\mu} < \mu_R$. □

Next we need to prove two lemmas. The first states that wages are lower in the convalescent region than at full employment. We need this result to show that prices will be higher in the convalescent region.

Lemma 8. *$\omega^*(\mu_t) < \omega^*_{\text{fe}}$ if $\mu_t \in (\underline{\mu}, \tilde{\mu})$.*

Proof. We know that $M(\mu_t) < \mu_t$ if $\mu_t \in (\underline{\mu}, \tilde{\mu})$.

$$\begin{aligned}
\omega^*(\mu_t) &= A - (\kappa + \chi\mu_t) + \beta(1-\delta)[\kappa + \chi M(\mu_t)] \\
&= A - \beta(1-\delta)\chi(\mu_t - M(\mu_t)) - (1 - \beta(1-\delta))(\kappa + \chi\mu_t) \\
&< A - (1 - \beta(1-\delta))(\kappa + \chi\mu_t) \\
&< A - (1 - \beta(1-\delta))(\kappa + \chi\underline{\mu}) = \omega^*_{\text{fe}}
\end{aligned}$$

□

Lemma 9. *Under Assumption 2, $\frac{W_t}{P_t} > (1-\delta)[\kappa + \chi\mu_t]$.*

Proof. We know that $\frac{W_t}{P_t} \geq \omega^*(\mu_t)$ by definition, so it suffices to show that $\omega^*(\mu_t) > (1-\delta)[\kappa + \chi\mu_t]$. In the flexible wage benchmark we have

$$\omega_t = \eta A + (1-\eta)b + \beta(1-\delta)q_{t+1}J_{t+1} \geq \eta A + (1-\eta)b > (1-\delta)(\kappa + \chi\mu_t)$$

for any $\mu_t \in [0,1]$, given assumption 2. □

Finally, we need to characterize dynamics of the economy starting at date 1, once the shock has abated. Under neutral monetary policy, if the ZLB never binds, allocations are (by definition) equal to those in the flexible wage benchmark. The following is immediate.

Lemma 10. *If $\mu_1 \geq \widetilde{\mu}$, the economy never returns to the full employment steady state.*

Proof. If the ZLB never binds, allocations are equivalent to those in the flexible wage benchmark, and we know that the economy never returns to steady state. It only remains to show that the ZLB can never help the economy converge to the full employment steady state. Suppose by contradiction that the economy converges to the full employment steady state. Let μ_t^R, μ_t^N denote allocations in the flexible wage benchmark and in the nominal economy, respectively, given the initial condition $\mu_1 \geq \widetilde{\mu}$. Let $T \geq 1$ be the first date at which $\mu_t^N < \mu_t^R$ (there must be some such date, since in the long run $\mu_t^N = 0, \mu_t^R > 0$, by assumption). Then we have

$$\begin{aligned}
J_{T-1}^N = \kappa + \chi\mu_{T-1}^N &= \kappa + \chi\mu_{T-1}^R = J_{T-1}^R \\
J_T^N = \kappa + \chi\mu_T^N &< \kappa + \chi\mu_T^R = J_T^R
\end{aligned}$$

This implies that real wages are higher at date $T-1$ in the flexible wage benchmark than in the nominal economy:

$$\begin{aligned} J_{T-1}^N &= J_{T-1}^R \\ A - \omega_T^N + \beta(1-\delta)J_T^N &= A - \omega_T^R + \beta(1-\delta)J_T^R \\ \omega_t^R - \omega_t^N &= \beta(1-\delta)(J_T^R - J_T^N) > 0 \end{aligned}$$

This is a contradiction - given the downward nominal wage rigidities, wages are always weakly higher than in the flexible wage benchmark. Thus the economy cannot converge to the full employment steady state. □

We are now ready to prove Proposition 4. Part 1. follows for the same reasons as in the previous lemmas. Define the function

$$B(\mu) = \frac{A-\kappa}{\omega(\mu) - (1-\delta)(\kappa + \chi\mu)}$$

on $(\underline{\mu}, \mu_R]$, where $\omega(\mu_1)$ denotes the prevailing real wage at date 1 as a function of μ_1. It is straightforward to show that $\omega(\mu_1)$ is continuous, and thus B is continuous. We have $B(\underline{\mu}) = \underline{\beta}$. Define $\overline{\beta} := B(\mu_R)$. If $\beta_0 > \overline{\beta}$, then if $\mu_1 = \mu_R$, we have

$$J_0 = A - \beta_0\omega(\mu_R) + \beta_0(1-\delta)(\kappa + \chi\mu_R) < \kappa$$

thus $\theta_0 = 0$, which is consistent with $\mu_1 = \mu_R$. If instead $\beta_0 \in (\underline{\beta}, \overline{\beta})$, then there exists $\mu \in (\underline{\mu}, \mu_R)$ such that $B(\mu) = \beta_0$, and a corresponding $\theta_0 = 1 - \frac{\mu_1}{1-(1-\delta)\mu_1}$. Then we have

$$J_0 = \kappa = A - \beta_0\omega(\mu_1) + \beta_0(1-\delta)(\kappa + \chi\mu_1)$$

and firms are indifferent between posting any number of vacancies; thus $\theta_0 \in [0,1]$ can indeed be an equilibrium. Finally, the fact that the economy does not return to full employment if it is thrown into the stagnant region follows from Lemma 10.

I Proof of Proposition 5

Under UMP, at date 0, the value of a filled vacancy J_0 is given by:

$$J_0 = A - \omega_0^{\text{ump}}(\beta_0) + \beta_0(1-\delta)J_{min} = \kappa$$

where we have used the definition of $\omega_0^{\text{ump}}(\beta_0)$. Since, $\mu_0 = 0$, the free entry condition is satisfied with $\theta_0 = 1$. This implies that $q_0 = 1$ and the economy remains at full-employment.Next, note that the date 0 Fisher equation implies:

$$1 + i_0 = \beta_0^{-1}\frac{P_1}{P_0} = 1$$

while at subsequent dates, we have:

$$1 + i_t = \beta^{-1} \frac{P_{t+1}}{P_t} = \beta^{-1} > 1$$

Thus, the ZLB constraint is satisfied at all dates (and binds only at date 0). The DNWR constraint is satisfied at all dates since $W_0 = W_{-1}$ and $W_0 > P_0 \omega^*(\mu_0)$, while for $t \geq 1$, $W_t \geq W_{t-1}$ and $W_t = P_t \omega^*(\mu_t)$.

J Wage Rigidities

J.1 Allowing for some nominal wage flexibility: $\varphi < 1$

Suppose that the downward nominal wage constraint is given by:

$$W_t \geq \varphi W_{t-1} \qquad \text{where} \qquad \varphi < 1$$

Compared to the original specification, this allows nominal wages to fall somewhat. With this wage constraint, the prevailing nominal wage will be:

$$W_t = \max \left\{ \varphi W_{t-1}, P_t \omega^* \left(\mu_t \right) \right\}$$

We will now show that the content of Proposition 4 remains true, i.e., under CMP:

1 The economy can only remain at the full employment steady state when $\beta_0 < \underline{\beta}$

2 When $\beta_0 \geq \overline{\beta}$, there is a hiring freeze at date 0, transitioning the economy to $\mu_1 = \mu_R$.

Proof of Claim 1 Suppose that if nominal wages can fall, the economy remains at the full employment steady state after a increase in β_0. Then it must be that $\frac{W_1}{P_1} = \omega^*_{\text{fe}}$. Since the ZLB does not bind at date 1, then monetary policy implements $W_1 = W_0$. Thus, we also know that $\frac{W_0}{P_1} = \omega^*_{\text{fe}}$. Since the ZLB binds at date 0 since $\beta_0 > 1$, we have:

$$1 = \beta_0 \frac{P_0}{P_1}$$

Using the fact that $\frac{W_0}{P_1} = \omega^*_{\text{fe}}$ in the Euler equation:

$$\frac{W_0}{P_0} = \omega_0 = \beta_0 \omega^*_{\text{fe}}$$

This is exactly the same expression as in Appendix G. Using this in the expression for J_0:

$$J_0 = A - \frac{W_0}{P_0} + \beta_0 \left(1 - \delta\right) J_{min} = A - \beta_0 \left[\omega^*_{\text{fe}} - \left(1 - \delta\right) J_{min} \right]$$

In Appendix G we proved that $\omega^*_{\text{fe}} - \left(1 - \delta\right) J_{min} > 0$. So J_0 is decreasing in β_0 and $J_0 = \kappa$ when $\beta_0 = \underline{\beta} := \frac{A - \kappa}{\omega^*_{\text{fe}} - (1-\delta) J_{min}}$ which is independent of φ. Thus, if $\beta_0 > \underline{\beta}$, then the economy cannot be at full employment at date 1.

Proof of Claim 2 Next, we show that for sufficiently high β_0, there is a hiring freeze at date 0. Notice that the value of a firm at date 0 is:

$$J_0 = A - \frac{W_0}{P_0} + \beta_0 (1 - \delta) (\kappa + \chi \mu_R)$$

where we are assuming that at date 1, the fraction of unskilled workers is μ_R. Since the ZLB binds at date 0, the bond Euler equation implies $1 = \beta_0 \frac{P_0}{P_1}$. Multiply both sides by W_0/P_0:

$$\frac{W_0}{P_0} = \beta_0 \frac{W_0}{P_1} = \beta_0 \frac{W_1}{P_1} \frac{W_0}{W_1} = \beta_0 \omega (\mu_1) \frac{W_0}{W_1}$$

Next, we establish that $W_0 \geq W_1$. We also know that $W_1 = \max \{\varphi W_0, P_1 \omega^* (\mu_1)\}$. Suppose it is the case that $W_1 = \varphi W_0$. This implies that $W_0 > W_1$. If instead $W_1 = P_1 \omega^* (\mu_1)$, we know from the definition of CMP that $W_0 \geq P_1 \omega^* (\mu_1) = W_1$ and so in either case, we have $W_1 \leq W_0$, i.e., nominal wages do not rise between date 0 and 1. Using the fact that $W_1 \leq W_0$ in the expression above:

$$\frac{W_0}{P_0} = \beta_0 \omega (\mu_1) \frac{W_0}{W_1} \geq \beta_0 \omega (\mu_1)$$

So, we know that:

$$J_0 = A - \frac{W_0}{P_0} + \beta_0 (1 - \delta) (\kappa + \chi \mu_R) \leq A - \beta_0 [\omega (\mu_R) - (1 - \delta) (\kappa + \chi \mu_R)]$$

Lemma 9 in the paper shows that $\omega (\mu_R) - (1 - \delta) (\kappa + \chi \mu_R) > 0$ for any μ. Thus, we can find a $\overline{\beta}$ for which $J_0 = \kappa$, i.e., there is a hiring freeze at date 0:

$$\overline{\beta} = \frac{A - \kappa}{\omega (\mu_R) - (1 - \delta) (\kappa + \chi \mu_R)}$$

Notice that this expression is independent of φ and is the same as the expression for $\overline{\beta}$ in the main paper where $\varphi = 1$.

J.2 State-dependent φ

While the explanation in Section 4.1.1 was made in the context of a model with a fixed $\varphi \in (0, 1]$, the same logic applies if we allow for state-dependent wage flexibility. Intuitively, we have shown for any $\varphi \in (0, 1]$, that real wages do not fall at date 0 even when nominal wages are allowed to adjust slightly downwards. Thus, for any $\varphi(\beta) \in (0, 1]$ or $\varphi(u) \in (0, 1]$, the same arguments follow through.

To see this, suppose the nominal wage constraint is given by:

$$W_t \geq \varphi (\beta_t) W_{t-1}$$

where $0 < \varphi (\beta_h) < \varphi (\beta_l) \leq 1$ for $\beta_h > \beta_l$ (the baseline model assumes $\varphi(\beta) = 1 \ \forall \beta$). In this case, nominal wages indeed fall by more following a larger demand shock. In partial equilibrium, this would reduce real wages and sustain hiring. However, in general equilibrium, the fall in nominal wages is also accompanied by a fall in prices, preventing real wages from falling. Thus, the content of Proposition 4 remains true in the context of this different downward nominal wage constraint. The cutoff points $\underline{\beta}$ and $\overline{\beta}$ are unaffected by the $\varphi(\beta)$ function. However, the behavior of nominal variables is different. A larger β_0 causes not only a larger increase in μ_1 as before but also a larger decline in W_0 to $\varphi(\beta_0) W_{-1} < W_{-1}$, and a larger decline in the price level P_0 to

$$P_0 = \frac{\varphi(\beta_0) W_{-1}}{\beta_0 \omega(\mu_1)}$$

State-dependent nominal wage flexibility does not prevent demand shocks from leading to unemployment. A larger fall in nominal wages simply results in a larger fall in prices, as households expect monetary policy to expect a lower price level at date 1 and curtail their consumption demand at date 0. Even if nominal wages are extremely flexible following large shocks, this nominal flexibility does not prevent equilibrium real wages from rising. There is no way to write a state-contingent rule for nominal wages which depends on the current state of the economy and prevents real wages from increasing in response to a demand shock, given CMP. However much nominal wages fall, prices fall to undo the effect on real wages.

K Assumptions on Search

K.1 Testing

The increasing relationship between the fraction of unskilled job-seekers and the expected cost of job creation is one of the key forces generating multiple steady states in our model. As we explain next, this increasing relationship continues to hold even if we allow firms to test applicants or observe their skill levels prior to hiring.

Case of perfect and costless testing Consider a test that fully reveals a worker's skill level to a firm at the time of meeting. Firms, however, still cannot prevent workers of different skill levels from contacting their vacancy. Further consider a firm that has posted a vacancy when the labor market is in the *slack* regime, i.e. $\theta < 1$. Given $m_t = \min\{l_t, v_t\}$, a vacancy meets an applicant with probability 1 when $\theta < 1$. Prior to training and bargaining, suppose that the firm observes the worker's skill level and has the option of turning the worker away. The firm gets zero payoff if it rejects a worker. Then, conditional on meeting a job-seeker, the firm hires the high-skill applicant only if $J_t \geq 0$ and hires the low-skilled applicant only if $J_t \geq \chi$. Given that $J_t \geq J_{min} = \frac{(1-\eta)(A-b)}{1-\beta(1-\delta)(1-\eta)}$, so as long as parameters are such that $J_{min} > \chi$, firms will never choose to reject either applicant. This condition on parameters is not ruled out by our assumptions. Thus, allowing the firms to test and fully observe the applicant's skill before they are trained and hired would not prevent the existence of multiple steady states.

More generally, while $J_{min} \geq \chi$ is a sufficient condition for multiple steady states to exist, multiple steady states can exist under weaker conditions. For example, multiple steady states can exist even if $J_{min} < \chi$ but the value of a filled vacancy at $\overline{\mu}$ (the high unemployment steady state) is greater than the training cost: $J_{ss} \equiv \kappa + \chi\overline{\mu} \geq \chi$ (which, again, does not violate the assumptions in the paper). This corresponds to a case in which the high unemployment steady state features such low wages that the value of a filled vacancy more than covers the cost of training an unskilled worker.

In order to generate multiple steady states, we do *not* need to assume that training unskilled workers is so costly that firms would turn these workers away, rather than train and hire them. All we need is that the value of a filled vacancy, net of training costs, is *lower* for an unskilled worker than for a skilled worker - this net value need not be *negative*. As long as the net value of an unskilled worker is less than the net value of a skilled worker, firms will be less willing to post vacancies when they anticipate meeting an unskilled worker with higher probability - even if ex post they would be willing to hire and train an unskilled worker. The benefit of posting a vacancy is $q_t(1-\mu_t)J_t + q_t\mu_t\,(J_t - \chi)$. With probability q_t, the firm meets an applicant. With probability $1-\mu_t$, the applicant is a skilled worker, in which case the value of filling the vacancy is J_t. With probability μ_t, however, the applicant is unskilled and the firm needs to train the worker before employing her, in which case the value of filling the vacancy is $J_t - \chi$. In both these cases, the firm's payoff ex post is positive, even though the firm has to incur the training cost for the unskilled worker (as long as $J_{ss} > \chi$). However, expected payoffs are lower *ex ante* whenever μ_t is higher. Our analysis with fully revelatory tests makes clear that the force generating multiple steady states in our model is firms' *ex-ante* uncertainty about applicants' skills prior to meeting, and not about the firm's uncertainty about the worker's skill level at the time of meeting.

Case of imperfect testing Suppose instead that parameters are such that firms *would* use test results in equilibrium, i.e. $J_{ss} < \chi$ and firms would reject an applicant who they know for sure to be unskilled.[24] Further assume that tests do not fully reveal an applicant's skill, that is, there is imperfect testing. The fact that firms *do* test workers' skills in equilibrium does not generally rule out multiple steady states. To see this, suppose that firms can costlessly conduct a test that reveals an applicant's skill with probability $p \in (1/2, 1)$ and misreports their skill with probability $1-p$. A firm can make its hiring decision contingent on the result of the test; if they hire an applicant who turns out to be unskilled, they train and employ that worker as in our baseline model.

An applicant who fails the test is unskilled with probability $\frac{p\mu}{p\mu+(1-p)(1-\mu)}$. Thus the firm will turn away applicants who fail the tests if

$$J < \chi \frac{p\mu}{p\mu + (1-p)(1-\mu)} \tag{24}$$

For relatively high κ, multiple steady states can exist as firms may not even turn away applicants who they *know* to be unskilled. Intuitively, a high value of κ allows ex ante uncertainty about the skill level

[24] If $J_{ss} \geq \chi$, as just described, firms would not turn away unskilled applicants even if they could perfectly identify them. Naturally then, in this case firms would not turn down applicants who fail an imperfect test, and multiple steady states would still obtain with imperfect testing.

of job applicants to depress hiring even when training costs are not too severe, that is, training costs are low enough such that unskilled applicants are trained and hired conditional on meeting. Consider now the opposite case, when κ is small (formally, $\kappa \to 0$). In this limit, applicants who fail a test are rejected in any interior steady state, since

$$J = \chi\mu < \chi\frac{p\mu}{p\mu + (1-p)(1-\mu)} \text{ for } \mu \in (0,1)$$

In a steady state in which failing applicants are rejected, the effective job finding rate for unskilled applicants is $(1-p)q < q$: they only find a job if they match with a firm (with probability q) and the test erroneously reports that they are skilled (with probability $1-p$). This worsens workers' outside option, lowering wages and increasing the value of a filled vacancy. The effective cost of job creation also changes relative to a steady state in which tests are not used. Testing has both costs and benefits. The benefit is that the firm economizes on training costs since it only trains those unskilled applicants who pass the test. The cost associated with using imperfect testing is that sometimes firms erroneously reject a skilled worker who fails the test. This means that more vacancies must be posted to create each job in expectation. The firm will only turn away applicants when the benefit of testing outweighs the cost, so the effective cost of job creation is lower when tests are used than when they are not used. When κ is small, the benefits tend to outweigh the costs and tests are more likely to be used.

Interior steady states are now described by the equation:

$$\underbrace{\frac{\kappa + (1-p)\,\chi\mu}{(1-\mu)\,p + \mu\,(1-p)}}_{\text{effective cost of job creation}} = \underbrace{\frac{(1-\eta)(A-b)}{1-\beta\,(1-\delta)\left[1-\eta\,(1-p)\,\frac{(1-\mu)}{(1-p)\mu+(1-\mu)p}\right]}}_{\text{value of job}} \tag{25}$$

Even when κ is small, if $p = 1/2$, tests are uninformative and are not used in equilibrium, as in our baseline model where multiple steady states exist. As p increases from $1/2$, firms are more likely to use tests in equilibrium. Even when tests are used in equilibrium, multiple steady states can exist so long as tests are not too informative. If κ is small, a sufficient condition ensuring such multiple steady states exist for sufficiently large χ is:

$$p \equiv \overline{p} < \frac{\beta(1-\delta)\eta}{1-\beta(1-\delta)(1-\eta)}$$

For example, with $\beta = 0.993$ and $\delta = 0.03, \eta = 0.7$ (Shimer, 2005), this cutoff equals 0.95, i.e. multiple steady states exist as long as tests are less than 95 percent accurate.

Figure 5 depicts this graphically, extending Figure 1 in the paper to the case where testing is possible and is used in equilibrium. The solid blue curve depicts the LHS of (25) (the effective cost of job creation), the solid red curve depicts the RHS (the value of a job), as functions of μ.

The dashed blue and red curves depict the corresponding functions in the baseline model without testing. The white area is the region in which (24) is satisfied and applicants who fail the test are rejected; the gray area is the region in which they are not rejected. Introducing informative tests shifts up the value of a job (moving from the dashed to the solid red curve). The effect of testing on the

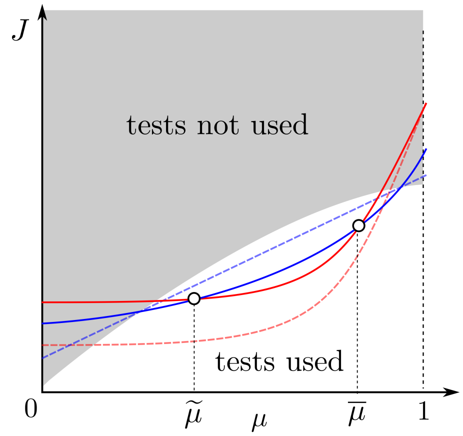


Figure 5. Multiple steady states with testing: The blue curve depicts the LHS of (25), the red curve depicts the RHS. The dashed blue and red curves depict the corresponding functions in the baseline model without testing. The white area is the region in which (24) is satisfied and applicants who fail are turned away; the gray area is the region in which they are not turned away.

cost of job creation - the difference between the dashed and solid blue curves - depends on μ. For very high or low μ, there is little ex ante uncertainty about an applicant's skill, and the benefit of testing is small and outweighed by the cost, so testing is not used. For intermediate values of μ, testing is used as it reduces the effective cost of job creation, and the solid blue line lies below the dashed blue line. When tests are not too informative ($p < \overline{p}$) and κ is small, the solid red and blue curves intersect in the white region, implying that even if testing is used in equilibrium, there are still three steady states: $\mu = 0$ (full employment), an unstable interior steady state $\widetilde{\mu} > 0$ and a high unemployment steady state $\overline{\mu}$.

So far we have considered the best case for testing to prevent the existence of multiple steady states, where it is costless to test. Even in this case, multiple steady states can emerge. If testing was very costly, it would not be used by firms in equilibrium and the results concerning multiplicity of steady states in our baseline model would continue to hold. However, if testing was costly but still used in equilibrium, firms would only pay the testing cost when the benefit of testing is sufficiently high, i.e., when there is a sizable fraction of unskilled job-seekers (high enough μ). Thus, it would still be the case that the expected cost of job-creation is increasing in μ, which is the main force which generates multiple steady states.

K.2 Directed Search and Segmented Markets

Our baseline model assumes random search. However, relaxing the assumption of random search does not necessarily eliminate multiple steady states, as we argue next.

Assume instead that firms can observe the skill of a job applicant and can post a vacancy specifying a particular skill level. Workers can choose to direct their search to any vacancy they choose. We continue to assume that firms must train unskilled workers if they hire them and wages are determined by Nash-bargaining. Whether introducing directed search in this way prevents multiple steady states from existing depends on two things: (i) whether unskilled applicants are profitable for a firm to hire

($J_{ss} > \chi$), as discussed above and (ii) whether firms can commit to turn away applicants who do not meet the particular skill requirement as specified in the vacancy posted.

Suppose that in equilibrium, there are two segmented markets - one where vacancies advertise for skilled applicants and one for unskilled applicants. Consider an unskilled worker who applies for a skilled job and matches with a firm. Turning the applicant away would waste the vacancy and yield a payoff of 0. Instead, training and hiring this worker yields a positive payoff for the firm if $J_{ss} > \chi$. Thus, a firm without commitment would not turn this worker away. Knowing this, unskilled workers would apply to this market and the equilibrium with segmented markets unravels. The only surviving equilibrium is the one described in the baseline model in which all workers and workers search in the same market. Thus, even if workers and firms are able to direct their search, markets may not be segmented in equilibrium and multiple steady states may exist. If firms can commit to turning away profitable workers, however, segmented markets would obtain, eliminating multiple steady states.

The discussion above assumes that firms can observe workers' skill level and showed that even under such a setting, directed search may not lead to segmented markets in equilibrium. Notably, it is even harder for segmented markets to emerge in equilibrium if workers' skills are not observable to firms. Under our maintained assumption that wages are Nash-bargained and firms cannot test workers, there is simply no way that firms can incentivize unskilled workers to reveal their types, i.e. apply in different markets. Even if we relax the Nash bargaining assumption and allow firms to write richer wage contracts, it is not trivial for firms to design such contracts so that applicants reveal their skill level. This is because skilled and unskilled workers have the same preferences over sequences of wages, so standard single-crossing conditions do not apply. Finally, if the firm is somehow able to write contracts which incentivize workers to apply in separate markets, these contracts involve some loss of profit for the firm because workers must be paid some information rent to induce them to reveal their types. If the information rent is very large, the firm may not not find it profitable to incentivize workers to reveal their type. Instead, it may be cheaper to offer the same contract to both types and simply train the unskilled applicants, leading to broadly the same outcomes as in our baseline model. Even if firms do use the separating contracts, they will only bear the cost of doing so if the fraction of 'bad' types μ is sufficiently high (in the extreme case, $\mu = 0$, there is no point in giving up information rent to screen workers as all workers are 'good') Thus, it would still be the case that the expected cost of job creation is increasing in μ, which is the force which makes multiple steady states more likely.

L Hysteresis since the Great Recession

We now ask whether the mechanisms described in the paper can help quantitatively explain the sluggish economic recovery following the Great Recession.

While the particular form of the matching function assumed above facilitates analytical results, it has the counterfactual prediction that the "high pressure" steady state has 0 percent unemployment. In what follows, we use a more standard matching function $m(v, l) = vl / \left(v^{\iota} + l^{\iota}\right)^{\frac{1}{\iota}}$. This allows us to consider a high pressure economy with an empirically plausible unemployment rate.

We calibrate the model to the U.S. economy. In our model, unemployed workers lose skill after one period. A period is six months, so "unskilled" workers correspond to those unemployed for 27 weeks or more, i.e., the long-term unemployed. Using resume audit studies, Kroft et al. (2013) find that the likelihood of receiving a call-back declines significantly after an unemployment spell, with most of the decline occurring within the first 8 months. Similarly, Ghayad (2014) finds a sharp decline in the call-back rate after 6 months. It therefore seems reasonable to posit that most of the skill loss upon losing a job occurs within the first 6 months.[25]

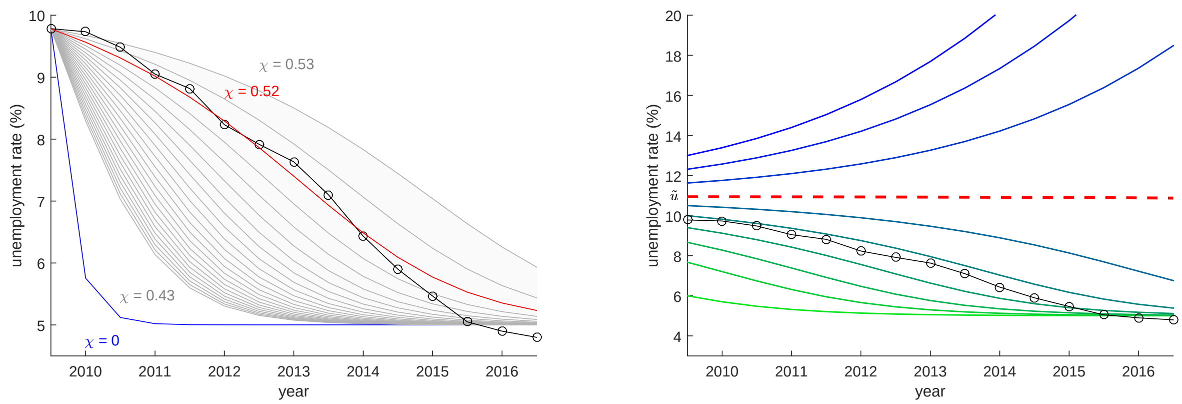
We set $\beta = 0.98$, implying a 4 % annualized steady state real interest rate. A is normalized to 1. We set $\iota = 0.5$ following Menzio and Shi (2011) and $\eta = 0.7$ (Shimer, 2005). b is chosen to imply a steady state replacement ratio of 70 percent (Hall, 2009). We set $\delta = 0.2105$ so that the 5 percent steady state unemployment is consistent with 20 percent of job seekers being long-term unemployed in steady state as observed in the U.S. before 2008 ($\mu_{ss} = 0.2$). The remaining moment - the steady state unemployment rate of 5% - pins down the value of a filled vacancy $J_{ss} = \kappa/f(\theta_{ss}) + \chi\mu_{ss}$. In what follows, we consider a range of values for (κ, χ) which satisfy this relation. We do *not* restrict attention only to combinations of (κ, χ) which result in multiple steady states.

Following the Great Recession, U.S. unemployment peaked at 10 percent in the second half of 2009 before beginning a slow decline, only returning to 5 percent after 6 years, as shown by the black line in Figure 6a. Our model can generate this initial rise in unemployment through a sequence of discount factor shocks which cause the ZLB to bind, as we discuss shortly. Can it also account for the subsequent path of unemployment? We have seen analytically (Proposition 2) that our model can in principle generate an arbitrarily slow recovery. To see whether our model can generate a slow recovery in a quantitatively plausible way without resorting to persistent shocks, we solve our model for a range of (κ, χ) combinations, starting from the unemployment rate of 10% observed in the second half of 2009. We implicitly assume that some sequence of β shocks brought the economy to this point, but these shocks have now abated and β is at the steady state level. For now, we do not specify the sequence of shocks that brought the economy to this level of unemployment since this does not affect dynamics of the recovery. The gray lines in Figure 6a show trajectories of unemployment for a range of (κ, χ) combinations consistent with multiple interior steady states, one of which is always 5 percent. Moving outwards from the origin, as we increase χ and decrease κ, the forces generating multiplicity become stronger and the recovery becomes slower. Quantitatively, the model can match the sluggish recovery observed in the data. The red line in Figure 6a indicates our preferred calibration, $\chi = 0.52$, which fits the data most closely. The implied κ is 0.0009.[26] While direct evidence on training costs

[25] Strictly speaking, in our simple model, skilled and unskilled job-seekers have the same job-finding rates and are paid the same wages conditional on employment: skill depreciation only shows up in the training costs faced by employers. For our purposes, all that matters is that the findings of these studies is consistent with a bulk of the skill depreciation occurring within 6 months of job-loss. Our rate of skill depreciation is also consistent with Ljungqvist and Sargent (1998); their calibration implies a 95% probability of losing skill after a 6 month unemployment spell. At the same time, setting a period to 6 months allows us to maintain the assumption that unemployed workers lose skill after one period, implying a one-to-one mapping between μ_t and u_{t-1}, allowing us to economize on state variables. If instead we assume that unemployed workers lose skill probabilistically, then we could reduce the length of a period but would have to deal with two endogenous state variables, complicating the analysis of our model due to the multiplicity of steady states. However, note that since separated workers can search for jobs within the same period, we are not imposing that any unemployed spell lasts at least half a year.

[26] This value of κ may seen small relative to some of the literature; this is because in our environment firms face an additional cost of creating a job (average training costs $\chi\mu$), so a smaller vacancy posting cost κ is needed to rationalize

is hard to come by, this value of χ lies well within the empirical estimates found in the literature.[27] Importantly, it is a result and not an assumption that our preferred χ lies within the parameter space consistent with multiple steady states. In this sense multiplicity is essential to match the sluggish recovery of unemployment.



(a) Model vs data **(b)** Different starting points

Figure 6. Trajectories of unemployment Panel (a) - black line:observed path of unemployment rate, grey lines: model trajectories of unemployment rate for various (κ, χ) combinations consistent with multiple steady states, red line: preferred calibration $\chi = 0.52$, blue line: model trajectory with $\chi = 0$. Panel (b) - solid lines denotes model trajectories of unemployment under preferred calibration given different values of initial unemployment rate. Dashed-red line: cutoff dividing regions of slow-recovery and permanent stagnation.

Under our preferred calibration, the sequence of shocks which generate a 5 percentage point increase in the unemployment rate is not implausibly large. While our analytical results considered only a one period shock, we now calibrate a persistent increase in β that lasts from the first half of 2008 to the second half of 2009 (3 periods) to match the observed rise in unemployment rates. This requires a sustained increase of β to 1.0173 for these 3 periods (corresponding to a -3.4% annual real interest rate), causing the ZLB to bind. These shocks raise real wages by 4.6% in the model. Empirically, real average hourly earnings[28] increased by 3.4%-3.5% over the corresponding period. As in the analytical model, neither the ZLB nor downward nominal wage rigidity bind after the shock abates.

Absent subsequent shocks, our mechanism is not just sufficient but necessary to explain the slow recovery. The blue line in Figure 6a shows the trajectory of unemployment when $\chi = 0$, i.e. a model

the observed level of unemployment.

[27] Barron et al. (1989) find that, on average, a new hire spends 151 hours on training in the first 3 months of the job. If only unskilled workers require training, as assumed in our model, this implies an upper bound of $\chi = 151/(0.2 * 1043.5) = 0.72$ (since the average fraction of unskilled workers in the US prior to 2008 was 20 percent, and assuming 2087 hour work-year, as is standard. Barron et al. (1989) also find that the median worker spends 81 hours in training. If we instead calibrate χ to match the difference between training costs between unskilled and skilled (median) worker, we get $\chi = (151 - 81)/(0.2 * 1043.5) = 0.34$.) The American Society for Training and Development (Paradise, 2009) estimated the average annual learning expenditure to be 2.24% of total annual payroll in 2008. In our model total training expenditures equal $\chi\mu\delta(1-u)$ in steady state while payroll equals $w(1-u)$ implying $\chi = 0.48$. Our preferred value of $\chi = 0.52$ lies comfortably within this range.

[28] We take the Average Hourly Earnings of all Private Employees from the Current Employment Statistics as our measure of nominal wages. We deflate this series using the Consumer Price Index for all Urban Consumers or Personal Consumption Expenditures Price Index, and compute the change in real hourly earnings between January 2008 and July 2009.

without training costs - essentially the standard DMP model, which has a unique steady state. Absent any further shocks, this model predicts a rapid recovery with unemployment returning to 5 percent by the end of 2010. This suggests that in the absence of persistent shocks, our mechanism is *necessary* to match the sluggish recovery observed in the data. As highlighted in Pissarides (2009), when firms post fewer vacancies due to poor aggregate conditions, competition for workers amongst recruiters declines, shortening the average duration of a vacancy. When the only costs associated with job creation are vacancy posting costs, the decline in vacancy duration lowers the average effective cost of job creation, mitigating the recession's adverse impact on hiring. Training costs, χ, undo this phenomenon, creating a protracted recovery. Unlike the average cost of vacancy creation, κ/f, which is pro-cyclical and rises when there is more competition amongst recruiters, training costs are counter-cyclical and rise when the composition of job-seekers tilts towards the unskilled.

Figure 6a suggests that the shock that hit the U.S. during the Great Recession was large enough to cause a slow recovery but not large enough to create permanent stagnation. The model allows us to evaluate how the economy would have responded had shocks been larger. Figure 6b shows the trajectory of unemployment under our preferred calibration given different initial unemployment rates (in the second half of 2009). Again, the black line plots data. The green lines show trajectories starting from lower initial unemployment; light blue lines indicate trajectories starting from higher unemployment. The red-dashed line shows the trajectory starting from $\tilde{u} = 10.9\%$ (the unstable steady state) which divides the regions of slow recovery and permanent stagnation. The figure shows that, had monetary policy been more accommodative after the initial shocks and kept the initial rise in unemployment below 8 percent, unemployment would have returned to 6% two years earlier, in 2012. More strikingly, had shocks been larger, causing unemployment to rise to 12%, the economy would have been unable to return to full-employment (absent fiscal policy).

Figure 6b suggests that had the US economy been hit by an even larger shock in 2007, it could have fallen into permanent stagnation absent a more timely and accommodative monetary policy response. In this regard, Europe presents a cautionary tale.[29] Figure 7 shows the fraction of long-term unemployed in Ireland, Greece, Spain, the Euro area and the U.S. from 2008 to 2016. While the fraction of long-term unemployed increased in the U.S. following 2007, timely monetary policy accommodation ensured that this increase was temporary. In contrast, the fraction of long-term unemployed increased following the European recessions of 2008 and 2011 and has since remained elevated.[30] Many commentators[31] have argued that the European Central Bank's response was insufficient from the point of view of these economies or came too late. The model suggests that this delayed or insufficient monetary policy response could explain why long-term unemployment has remained persistently high in these economies. From this perspective, ECB President Draghi's

[29] Since the assumption of risk neutrality in our model implies that real interest rates are exogenous, the analysis in the model can also be applied to a small open economy in a monetary union.

[30] Of course, experiences differed widely across European countries; for example, the fraction of long -term unemployed actually declined in Germany over this period. Here we focus on those countries most severely affected during the European crisis, from whose perspective the ECB's monetary policy was arguably insufficiently accommodative. Having said that, despite the heterogeneous experiences of different countries, on average, the Euro area did experience an increase in the fraction of long-term unemployed.

[31] See for example Kang et al. (2015).

"whatever it takes" speech in July 2012 - which can be seen as a commitment to very accommodative policy - may have come too late to reverse the effects of hysteresis. Thus, to prevent hysteresis, monetary stimulus must not just be large, but also timely.

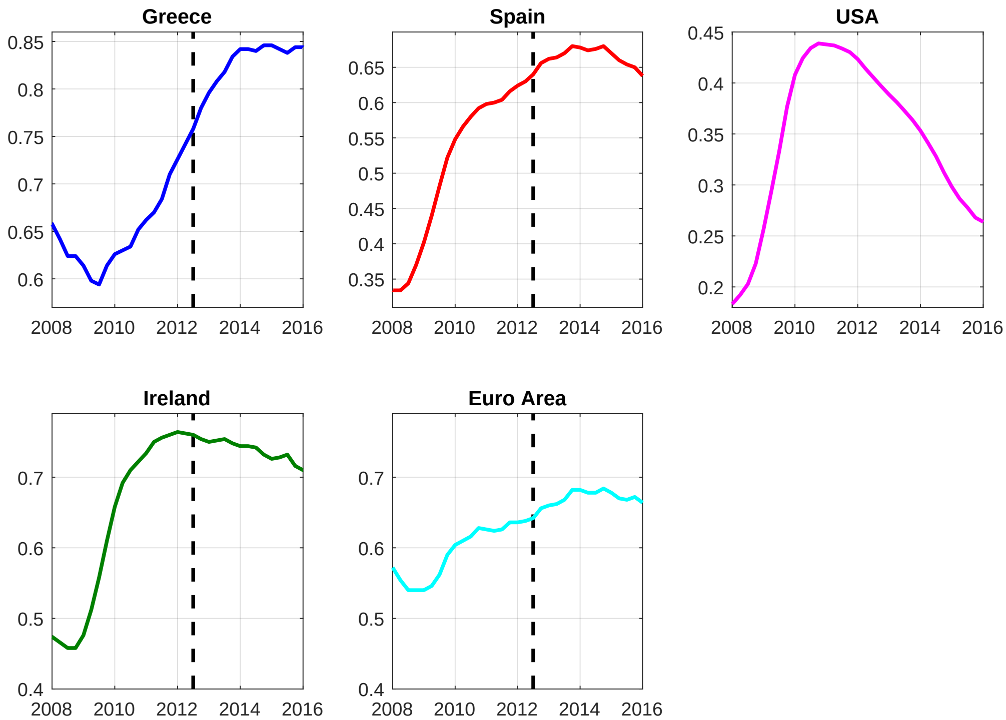
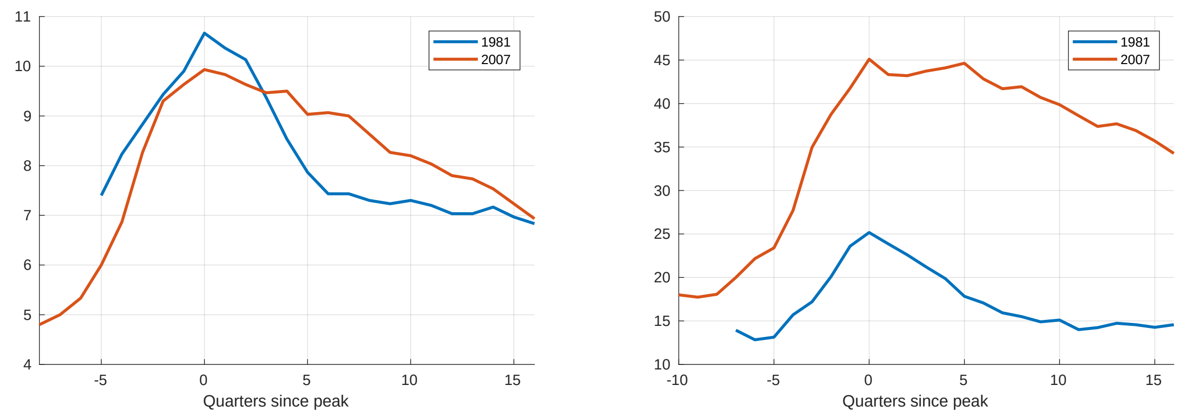


Figure 7. Fraction of long-term unemployed (>27 weeks) in select countries. The figure plots five quarter moving averages of quarterly data. The dashed-line indicates the timing of Draghi's "whatever it takes" speech. Source: Eurostat and FRED.

The Great Recession vs. the Volcker Recession Our analysis predicts that an economy vulnerable to hysteresis will recover quickly following a small increase in unemployment but will take much longer to recover following a large increase in unemployment. Through the lens of our model, the recovery from the Great Recession was so protracted because the initial increase in unemployment was so large compared to the previous few recessions. However, the 1981 recession saw unemployment rates increase to an even higher level (10.8%) than in the Great Recession (10%). Figure 8a shows the behavior of the unemployment rate around its peak (normalized to date 0) following the recessions beginning in 1981Q3 and 2007Q4. While unemployment rates returned to its pre-recession levels 5 quarters after reaching its peak after the 1981 recession, unemployment rates only returned to their pre-recession level 6 years after the 2007 recession.

This might appear inconsistent with our model's predictions. Recall however, that the key variable for the mechanism in the model is not the unemployment rate itself but the fraction of unskilled job-seekers. Empirically, this fraction can be proxied by the fraction of unemployed individuals who have been unemployed for more than 27 weeks, i.e. the long-term unemployed, given the assumption that workers are more likely to lose skill the longer they have been unemployed.[32] Figure 8b shows that the fraction of long-term unemployed individuals around its peak in the two recessions. This fraction increased much less in the 1981 recession compared to the 2007 recession, and accordingly returned to its pre-recession level much more rapidly. Through the lens of the model, a smaller increase in long-term unemployment begets less skill depreciation (a smaller increase in μ), and a speedier recovery. While the 1981 recession was comparable to 2007 in terms of peak unemployment, it was much smaller in terms of the fraction of long-term unemployed.



(a) Unemployment Rate **(b)** Fraction of long-term unemployed (>27 weeks)

Figure 8. Recoveries following the 1981 and 2007 recessions

The mapping between the overall level of unemployment and the fraction of long-term unemployed might have changed over the last five decades due to slow moving structural changes in the labor market which are outside the scope of this paper. For example, factors such as the trend decline in separation rates and the reduced importance of temporary layoffs relative to the 1980's would tend to make a given unemployment rate associated with a higher fraction of long-term unemployed.

[32]In our stylized model, this proxy is exact since unemployed workers become unskilled after 1 period, i.e. 6 months. More generally, the two measure there would be positively correlated rather than identical if workers lost skill probabilistically conditional on being unemployed.