Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy

Giorgio E. Primiceri*
Princeton University

First draft: July 2003
This version: May 2004

Abstract

This paper provides an explanation for the run-up of US inflation in the 1960s and 1970s and the sharp disinflation in the early 1980s, which standard macroeconomic models have difficulties in addressing. I present a model in which rational policymakers learn about the behavior of the economy in real time and set stabilization policy optimally, conditional on their current beliefs. The steady state associated with the model's self-confirming equilibrium is characterized by low inflation. However, prolonged and asymmetric episodes of high inflation can occur when policymakers underestimate *both* the non-accelerating inflation rate of unemployment *and* the persistence of inflation in the Phillips curve. I estimate the model using likelihood methods. The estimation results show that the model accounts remarkably well for the evolution of policymakers' beliefs, stabilization policy and the postwar behavior of inflation and unemployment in the United States.

*I am deeply indebted to Chris Sims and Mark Watson for their guidance and advice. I am also grateful to Jean-Philippe Laforte, Francesco Lippi, Guido Lorenzoni, Jonathan Parker, Bruce Preston, Lars Svensson, Lawrence Uren, Noah Williams, Michael Woodford, seminar partecipants at Princeton University, the Federal Reserve Bank of New York, University of Chicago, Northwestern University, Harvard University, University of Pennsylvania, University of Michigan, Columbia University, the Federal Reserve Board, University of California at San Diego for their comments, and especially to Alejandro Justiniano, Andrea Tambalotti and Thijs van Rens for countless discussions about my work. Remaining errors are my own.
Department of Economics, Princeton University, Princeton, NJ 08544, USA. Email: gprimice@princeton.edu. URL: http://www.princeton.edu/~gprimice.

1 Introduction

This paper aims to explain the behavior of inflation and unemployment in the United States. Figure 1 presents a plot of the annualized quarterly growth rate of the GDP deflator and the total civilian unemployment rate over the postwar period. The striking feature of the graph is the long and pronounced run-up of inflation, which occurred in the 60s and 70s. This episode, known as the Great Inflation, is not just "America's only peacetime inflation" (DeLong 1997), but has also been called "the greatest failure of American macroeconomic policy in the postwar period" (Mayer 1999).

The Great Inflation is characterized by at least four stylized facts.

- *Dimension.* Between 1963 and 1981 the inflation rate in the United States rose by more than 9 percentage points. If we exclude the peak in 1974 (which is due to the effect of the first oil price shock), the rate of increase was approximately constant.
- *Duration.* The episode of high inflation lasted for more than 20 years. Inflation started to increase around 1963 and came back under control, at a level of about 2%, only around 1985.
- *Asymmetry.* The episode of high inflation was asymmetric. In the early 80s, the duration of the so called "Volcker disinflation" was much shorter than the phase of rising inflation.
- *Unemployment lagged inflation.* Unemployment lagging behind inflation is a general characteristic of the business cycle. However, this feature of the data was particularly evident in the period of high inflation, with unemployment peaking always a few quarters after inflation.

This paper puts forward a theory of the behavior of inflation and unemployment, which fits the US data well and, in particular, explains all four of the stylized facts above. This theory is based on the evolution of policymakers' beliefs about the structure of the economy. Previous attempts to explain the Great Inflation fall apart in three categories, which I label the "bad luck," the "lack of commitment" and the "policy mistakes" views. I briefly discuss each of these branches of literature below.

The "bad luck" view The first type of explanations is based on bad luck. It has been well documented that the volatility of the exogenous non-policy shocks was higher in the 60s and 70s than in the last two decades of the century (see, for instance, Blanchard and Simon 2001, Cogley and Sargent 2003b, Kim and Nelson 1999a, Kim, Nelson, and Piger 2001, McConnell and Perez-Quiros 2000, Primiceri 2003, Sims and Zha 2004, Stock and Watson 2002, Stock and Watson 2003b). However, although non-policy shocks definitely played an important role, it is hard to reconcile the existing estimates with the dimension and the duration of the Great Inflation.

The "lack of commitment" view The second class of explanations is what Christiano and Fitzgerald (2003) have called the "institution vision of inflation". According to this view, inflation was high in the 60s and 70s because policymakers did not have any incentive to keep inflation low. The motivation for this relies on the time-consistency problem of optimal policy, first emphasized by Kydland and Prescott (1977) and Barro and Gordon (1983). If the unemployment rate targeted by policymakers is lower than the natural unemployment rate, policymakers try to get close to the target by creating inflation surprises. However, if private agents are rational, they recognize such opportunistic behavior and revise upwards their expected inflation. This introduces a positive inflationary bias in the economy (without pushing unemployment below its natural rate). In other words, inflation rose because of the lack of ability to commit to a low inflation regime. The importance of this line of research has been recently emphasized by Chari, Christiano, and Eichenbaum (1998), Christiano and Gust (2000) and Christiano and Fitzgerald (2003).

However, the inflation bias generated by the time-consistency problem seems to be quantitatively too small to explain the high inflation of the 60s and 70s (see, for example, Sargent 1999). Ireland (1999) formally tests the inflation bias hypothesis. While he is not able to reject it, his estimates suggest the presence of an inflationary bias of small magnitude.

Moreover, it is hard to reconcile the time-consistency view with the rapid Volcker disinflation. In fact, it is not clear what exactly changed between the pre and post 80s period from the institutional point of view.[1] The final difficulty with the "lack of commitment" approach is the fact that it would predict unemployment leading, rather than lagging inflation. This is due to the fact that the advantages of inflationary surprises depend on the level of unemployment. As mentioned above, this is clearly at odds with the data.

The "policy mistakes" view This approach focuses on policy mistakes and stresses that in the 60s and 70s monetary policymakers were not as good as the ones of the last two decades. For example, many authors have argued that US monetary policy was less responsive to inflationary pressures under the Fed chairmanship of Arthur Burns than under Paul Volcker and Alan Greenspan (among others, see Boivin and Giannoni 2002, Clarida, Gali, and Gertler 2000, Cogley and Sargent 2001, Cogley and Sargent 2003b, Judd and Rudebusch 1998, Lubik and Schorfheide 2004).[2] The problem with this branch of the literature is that often it does not provide any explanation why the policy authorities behaved so differently in the pre and post 80s period.

In this respect, the line of research started by Orphanides represents an attempt to rationalize the behavior of policymakers. Orphanides (2000 and 2002) has argued that policymakers in the 70s overlooked a break in potential output. They overestimated potential output leading to over-

[1] Recent work has made some progress in this direction. Sargent (1999), for example, explains the disinflation as escape dynamycs from the inflation biased equilibrium. Rogoff (2003) argues that Central Banks' lower incentive to inflate is related to globalization and the consequent increase in world competition. Albanesi, Chari, and Christiano (2003), instead, analyze the lack of commitment problem in an optimizing agents model and show the existence of multiple equilibria, which can potentially explain the disinflation. On the other hand, Albanesi, Chari, and Christiano (2001) analyze particular classes of models and show that in many cases there is no time-consistency problem at all.

[2] This view is controversial. Other studies have in fact found either little evidence of changes in the systematic part of monetary policy (for example, Bernanke and Mihov 1998, Hanson 2003, Leeper and Zha 2002, Primiceri 2003) or no evidence of unidirectional drifts in policy toward a more active behavior (Sims 1999, Sims 2001b, Sims and Zha 2002).

expansionary policies, which ultimately resulted in high inflation. Among others, this explanation has also been proposed by Lansing (2002), Bullard and Eusepi (2003), Reis (2003) and Tambalotti (2003). While this strand of literature represents a step forward, the dimension of the high inflation episodes explained by such models is usually much lower than what we actually observe in the data. This is the case both when policymakers are assumed to behave optimally (as in Reis 2003) and when they follow a non-optimal Taylor-type rule, like in most of the papers. Furthermore, the explanations based on the misperception of potential output fail to address the Volcker disinflation, unless an exogenous shift in policymakers' preferences is specified (see, for instance, Bullard and Eusepi 2003).

Contribution of the paper While there is clearly some truth in all of these theories, they also seem to have difficulties in addressing at least some of the stylized facts of the hump-shaped behavior of inflation and unemployment. This paper proposes instead an explanation of the Great Inflation that matches all these stylized facts.

I present a model, in which rational policymakers form their beliefs about the behavior of the economy in real time and set stabilization policy optimally, conditional on the information available to them. Hence, the main ingredient of the model is policymakers' adaptive learning about the structure of the economy, which determines the stance of aggregate demand policy.

Although the equilibrium of the model is characterized by low inflation, episodes of high inflation and unemployment can occur when policymakers simultaneously underestimate *both* the non-accelerating inflation rate of unemployment *and* the persistence of inflation in the Phillips curve. Such initial conditions result in peculiar dynamics of policymakers' beliefs, ultimately affecting also their perception of the slope of the Phillips curve and of the cost of the inflation-unemployment trade-off.

Intuitively, if real-time policymakers underestimate the non-accelerating inflation rate of unemployment, this results in overexpansionary policies and higher inflation. It is possible to show that optimizing policymakers who have the correct estimates of the coefficients of the Phillips curve would respond to inflation relatively promptly, preventing inflation from rising. However, if policymakers' estimate of the persistence of inflation in the Phillips curve is downward biased, they rationally choose not to react strongly to inflation, amplifying the initial effect. The reason is that the more stationary inflation is perceived to be, the sooner it is expected to revert to its mean and the less urgent is the need for anti-inflationary action. This period of "overoptimism" ends when inflation reaches a level that concerns policymakers. However, anti-inflationary policy is postponed even further because, when policymakers start reacting, pushing unemployment above the *perceived* non-accelerating inflation rate does not seem to reduce inflation. This is because they still have a downward biased estimate of the non-accelerating inflation rate of unemployment. In this period of "overpessimism", they temporarily and mistakenly perceive a very costly inflation-unemployment trade-off. When policymakers' beliefs finally converge to their equilibrium value and the perceived trade-off improves, inflation has already reached a very high level. As a result, the policy action is strong and decisive and the disinflation is sharp. In other words, the disinflation occurs when, the perceived inflation-unemployment trade-off becomes favorable, relative to the level of inflation.

From a qualitative perspective, I provide narrative evidence from the Economic Report of the President to demonstrate that the above is a realistic description of the evolution of US policymakers' beliefs in the 60s, 70s and 80s. Among others, Orphanides (2000), DeLong (1997) and Romer and Romer (2002) have argued in favor of the policy misperception of potential output and the non-accelerating inflation rate of unemployment in the 60s and 70s. Policymakers' misperception of the persistence of inflation in the Phillips curve in the 60s is also no longer controversial. For example, Blanchard and Fischer (1989) and Mayer (1999) have noted that, at least until the early 70s, most of the econometric studies underestimated inflation persistence. In relation to the over-pessimism phase, DeLong (1997) and Romer and Romer (2002) have provided narrative evidence that policy was cautious in the 70s because the cost of lowering inflation seemed too high.

From a quantitative and statistical standpoint, I show that the evolution of policymakers' beliefs about the coefficients of the Phillips curve is very important to explain the behavior of inflation and unemployment. I estimate the model using likelihood methods. The estimated version of the model accounts remarkably well for the evolution of policymakers' beliefs, stabilization policy and the postwar behavior of inflation and unemployment in the United States.

The importance of policymakers' learning dynamics has been recognized by many authors. In the context of the "natural rate" literature, policymakers' learning has been introduced by Sims (1988). Theoretical advances include Sargent (1999), Cho, Williams, and Sargent (2002) and Williams (2003). Empirical studies include Chung (1990) and Sargent (1999). The main insight of this literature is that policymakers' learning introduces temporary deviations from the model's equilibrium, which is characterized by an inflation bias. These temporary deviations are in the direction of the optimal, low inflation outcome. Unlike these studies, in this paper the time-consistency problem plays a minor role, and the equilibrium outcome is a low inflation regime. Nevertheless, the model explains the run-up of US inflation in the 60s and 70s and the sharp disinflation in the early 80s. Although the explanation roughly belongs to the "policy mistakes" category, in this paper policymakers are assumed to be rational and optimizing. As a consequence, I find that the mismeasurement of the non-accelerating inflation rate of unemployment alone is not sufficient to generate fluctuations of the inflation rate comparable to what we observe in the data. This differs importantly from Orphanides (2000) and other similar approaches.

The paper is organized as follows. Section 2 presents the benchmark theoretical model of the economy and policymakers' behavior. Section 3 illustrates the main model's implications using a simplified version of the benchmark model, for which it is possible to obtain some analytic results. The model is then used to provide an explanation of the Great Inflation. Since the explanation is based on the evolution of policymakers' beliefs, section 3 also provides narrative evidence in favor of this channel. Section 4 focuses instead on statistical evidence, i.e. estimation, fit and quantitative simulations results. Sections 5 and 6 demonstrate the robustness of the results to more sophisticated assumptions about the behavior of private agents and policymakers. In particular, section 5 introduces private agents' forward looking behavior in the model. Section 6 allows for stochastic volatility of the exogenous innovations and for more informed policymakers, knowing the stochastic process of the non-accelerating inflation rate of unemployment. Section 7 makes an attempt to uncover the deeper reasons of the Great Inflation, i.e. why policymakers underestimated the persistence of inflation in the Phillips curve in the 60s. In order to do so, I estimate a time varying version of the Phillips curve, in which both the persistence parameters and the disturbances variance are allowed to change over time following a Markov switching process. I find that the high volatility of temporary shocks during the 50s is the most important cause of the serious bias in the estimation of inflation persistence, which, coupled with the underestimation of the non-accelerating inflation rate of unemployment, resulted in very high inflation. Section 8 concludes with some final remarks.

2 Imperfect Information and Inflation-Unemployment Dynamics

In this section I present a simple model of inflation-unemployment dynamics when policymakers have imperfect information. The source of imperfect information is the fact that policymakers do not know the exact model of the economy. In particular, they are uncertain about the value of the model's parameters. Therefore, policymakers update their beliefs about the model's unknowns in every period and implement optimal policy, conditional on their current beliefs. In turn, the policy variable affects the behavior of inflation and unemployment because it enters the model describing the true evolution of key macroeconomic variables.

2.1 The Model Economy

While the case of a microfounded, forward looking model will be analyzed in section 5, as a benchmark I consider a simple rational expectations model that can be rewritten as a backward looking one. Even if conceptually similar to modern New-Keynesian specifications, the benchmark model is more in the spirit of the empirical literature following along the lines of Gordon (1982 and 1998), King, Stock, and Watson (1995) and, more recently, Staiger, Stock, and Watson (1997 and 2001) and Rudebusch and Svensson (1999).

The choice of this framework is motivated by three reasons. First, the backward looking specification is simple and transparent. Therefore it provides a clear and direct intuition for the role played by policymakers' learning dynamics in the behavior of inflation and unemployment. Second, the backward looking model fits the data well. Third, the backward looking model is tractable and convenient for estimation. Therefore, instead of arbitrarily calibrating the model, I can rely on the estimated parameters to perform simulations. Integrating learning dynamics in a forward looking specification is instead computationally very expensive and, in some cases, prohibitively so for estimation. I will return to this point in section 5.

The private sector part of the model is described by the following equations:

$$\pi_t = \pi_t^e - \tilde{\theta}(L)(u_{t-1} - u_{t-1}^N) + \varepsilon_t. \tag{1}$$

$$(u_t - u_t^N) = \rho(L)(u_{t-1} - u_{t-1}^N) + V_{t-1} + \eta_t, \tag{2}$$

$$u_t^N = (1-\gamma)u^* + \gamma u_{t-1}^N + \tau_t. \tag{3}$$

Equation (1) represents a Phillips curve, where π_t is the inflation rate, π_t^e is the agents' expected inflation rate, u_t is the unemployment rate and u_t^N is the time varying non-accelerating inflation rate of unemployment (NAIRU). $\tilde{\theta}(L)$ is a lag polynomial and ε_t is a random innovation, assumed to be *i.i.d.* $N(0, \sigma_\varepsilon^2)$.[3] I assume that some of the agents are fully rational, while the rest of them form their expectations adaptively, so that

$$\pi_t^e = (1 - \tilde{\alpha}(1))E_{t-1}\pi_t + \tilde{\alpha}(L)\pi_{t-1}. \tag{4}$$

$\tilde{\alpha}(L)$ is a lag polynomial and the combination of (1) and (4) leads to the following familiar reduced form Phillips curve:[4]

$$\pi_t = \alpha(L)\pi_{t-1} - \theta(L)(u_{t-1} - u_{t-1}^N) + \varepsilon_t, \tag{5}$$

where $\alpha(L) = \frac{\tilde{\alpha}(L)}{\tilde{\alpha}(1)}$ and $\theta(L) = \frac{\tilde{\theta}(L)}{\tilde{\alpha}(1)}$. Note that $\alpha(1) = 1$, implying the absence of a long run trade-off between unemployment and inflation, which is consistent with the natural rate hypothesis. The interpretation of (5) is straightforward: the inflation rate changes either because of a random "cost push" term or because unemployment is not in line with its non-accelerating inflation level.

Equation (2) is a very simple aggregate demand equation, where $\rho(L)$ is a lag polynomial, η_t is an *i.i.d.* $N(0, \sigma_\eta^2)$ random innovation, and V_t is a variable controlled by policymakers. In other words, unemployment deviates (persistently) from the NAIRU either because of a random shock or because of policymakers' decisions about stabilization policy.[5]

Notice that the policy variable V_t can be interpreted as capturing the joint effect of monetary and fiscal policy. In particular, this modeling strategy avoids complications related to the specification of two aspects of the policy process: the relative importance of the monetary and fiscal policy actions on real activity and the particular channels through which monetary and fiscal policy affect real activity. This is in the spirit of the recent renovated interest on the effect of fiscal policy (see, among others, Blanchard and Perotti 2002, Perotti 2002, Uhlig 2002) and the direct role of monetary aggregates in macro models (Favara and Giordani 2002, Leeper and Roush 2003).[6]

[3] The case of heteroskedastic innovations is particularly interesting in the context of learning models and will be analyzed in section 6.1.

[4] Among others, Gordon (1997 and 1998), Staiger Stock and Watson (1997 and 2001), Rudebusch and Svensson (1999) and, more recently, Onatski and Williams (2003) and Stock and Watson (2003a) have used specifications similar to (5).

[5] Specifications similar to (2) can be found in Fuhrer and Moore (1995b) and Rudebusch and Svensson (1999) and, more recently, Onatski and Williams (2003) and Stock and Watson (2003a).

[6] See, for instance, Romer and Romer (2003) for a summary of the problems of concentrating only on the federal funds rate as an indicator of policy. Primiceri (2003) provides empirical evidence that policy represented only by a short term interest rate is not able to account for the poor economic performance of the 70s in the US.

Equation (3) describes the exogenous stochastic process for the NAIRU, which is assumed to evolve as an AR(1), where τ_t is $i.i.d.$ $N(0,\sigma_\tau^2)$. u^* represents the unconditional expectation of u_t^N. This assumption on the evolution of u_t^N is standard in the literature (see, for instance, Staiger, Stock, and Watson 2001, although they set $\gamma = 1$ in their empirical specification).

2.2 Optimal Policy under Imperfect Information

The value of the policy variable V is chosen in every period by policymakers. They base their decision on the available information and on current beliefs about the state of the economy. I assume that policymakers know the structure of the true model of the economy (given by equations (5) and (2)), but they are uncertain about the value of the unobservable variables (the NAIRU) and the coefficients. Policymakers estimate the model's parameters in every period and use these estimates as true values, neglecting both estimates uncertainty and the possibility of future updates.[7]

Policymakers' beliefs about unobservables and the model's constant coefficients are denoted by hats. All these beliefs are formed at time t, but the subscript is omitted for simplicity. In particular, $\hat{u}_{t-1}^N$ stands for $\hat{u}_{t-1|t}^N$ and indicates the estimate at time t of the value of the NAIRU at time $t-1$. Policymakers determine the optimal value of the policy variable V by solving the following optimization problem:

$$\min_{\{V_s\}} L = \hat{E}\sum_{s=t}^{\infty}\delta^{s-t}\left[(\pi_t-\pi^*)^2+\lambda\left(u_t-k\hat{u}_t^N\right)^2+\phi\left(V_t-V_{t-1}\right)^2\right], \quad (6)$$

$$\text{s.t.} \quad \pi_t = \hat{c}_\pi+\hat{\alpha}(L)\pi_{t-1}-\hat{\theta}(L)(u_{t-1}-\hat{u}_{t-1}^N)+\hat{\varepsilon}_t, \quad (7)$$

$$(u_t-\hat{u}_t^N) = \hat{c}_u+\hat{\rho}(L)(u_{t-1}-\hat{u}_{t-1}^N)+V_{t-1}+\hat{\eta}_t. \quad (8)$$

L represents the familiar quadratic loss function, which depends on deviations of inflation and unemployment from the respective targets. λ represents the weight on the unemployment objective. Notice that, like in Barro and Gordon (1983), the target for the unemployment rate is given by $k\hat{u}_t^N$. When $k=0$ the unemployment target is equal to zero and policymakers' preferences resemble Kydland and Prescott's (1977). This would be the case in which the policy time-consistency problem is most pronounced. On the other hand, when $k=1$, the target is the non-accelerating inflation unemployment rate and the time-consistency and the related inflation bias completely disappear from the policy problem. Blinder (1998), among others, has argued in favor of these kind of policy preferences. The loss function has also a "smoothing" component, which penalizes big shifts of the policy variable. From an empirical perspective, this term is crucial in order to match the actual policy behavior because it helps to account for the strong autocorrelation shown by the instruments of economic policy (Dennis 2001, Favero and Rovelli 2003 and Soderstrom, Soderlind, and Vredin 2003). See also Woodford (2003) for an overview of the theoretical desirability of the smoothing term in the monetary policy context. Moreover, it easy to think to models in which instrument smoothing is desirable also for fiscal policy (for example, Barro 1979).

[7] These assumptions are standard in the adaptive learning literature, although the resulting policymakers' behavior is suboptimal. For alternative approaches, see Beck and Wieland (2002) and Wieland (2000a and 2000b).

Policymakers minimize their loss function subject to two constraints, (7) and (8). These constraints are the estimated counterparts of the true Phillips curve and aggregate demand equations. Observe that in the policymakers' model $\hat{\alpha}(1)$ is not constrained to be equal to one and $\hat{c}_\pi$ in (7) controls their beliefs about the level of average inflation. $\hat{c}_u$ in (8) instead controls their beliefs about the effect of setting V equal to zero. In other words, $-\hat{c}_u$ represents the "natural" level of policy, i.e. the level of V that does not affect unemployment.[8] Notice that, without further assumptions, $\hat{c}_\pi$, $\hat{c}_u$ and $\hat{u}_t^N$ would not be separately identified in the policy econometric model. A discussion about this issue is postponed until the next subsection.

The optimal rule for fixing V_t is given by

$$V_t = g(\hat{\beta}) S_t, \tag{9}$$

where $\hat{\beta}$ represents the vector of values for the model's parameters that policymakers treat as certainty-equivalents; $g(\hat{\beta})$ is the standard solution of a linear-quadratic problem, obtained solving the corresponding Riccati equation. S_t meanwhile denotes the set of relevant state variables and beliefs about unobservable states of the economy. To be more concrete, assuming that all the lag polynomials have order two, $\hat{\beta}$ would be given by the vector $\left[\hat{c}_\pi; \hat{\alpha}_1; \hat{\alpha}_2; \hat{\theta}_1; \hat{\theta}_2; \hat{c}_u; \hat{\rho}_1; \hat{\rho}_2; \hat{u}_t^N\right]$ and S_t by the vector $\left[1; \pi_t; \pi_{t-1}; u_t - \hat{u}_t^N; u_{t-1} - \hat{u}_{t-1}^N; V_{t-1}\right]$.

Once (9) is substituted into (2), the dynamics of inflation and unemployment are fully specified by the Phillips curve and the aggregate demand equation. This makes explicit the channel through which policymakers' beliefs affect the behavior of inflation and unemployment.

2.3 Learning

To implement the optimal rule and fix the value of the policy variable V_t, policymakers must estimate the parameters of interest, which are the unobservables and the coefficients. While I will relax this assumption in section 6.2, in most of the paper I will assume that policymakers form their beliefs about the NAIRU using univariate methods, i.e. they extract information about the NAIRU, only looking at the behavior of unemployment. Observe that this is suboptimal as, conditional on the true model of the economy, better estimates of the NAIRU could be obtained by exploiting the information contained not only in the unemployment rate, but also in the inflation rate. However, there are at least three reasons motivating this choice.

First, historical narrative evidence (see, for instance, DeLong 1997, Romer and Romer 2002) suggests that this is a realistic assumption for the behavior of past policymakers. Even now, univariate algorithms are commonly used to define the potential of the economy, especially in the output gap and monetary policy literature (see, for instance, Orphanides and Van Norden 2001, Lansing 2002, Taylor 1999).

Second, Staiger, Stock, and Watson (2001) show that the NAIRU estimated using formally the Phillips curve approach is basically indistinguishable from the univariate trend in unemployment.

[8] In the true model the "natural" level of policy is normalized to zero.

The third reason is substantial. In fact, as mentioned above, $\hat{c}_\pi$, $\hat{c}_u$ and $\hat{u}_t^N$ are clearly not separately identified in equations (7) and (8). In the true Phillips curve, (5), the sum of coefficients on lagged inflation is equal to one. Therefore, as long as inflation does not have a time trend, the constant in the Phillips curve must be equal to zero. On the contrary, policymakers do not have a model in which the sum of coefficients on lagged inflation is restricted to one. This means that they must estimate not only the coefficients on lagged inflation, $\hat{\alpha}(L)$, but also the constant of equation (7), which captures their beliefs about average inflation.[9] So it is clear that the estimation of (7) cannot identify $\hat{c}_\pi$ from $\hat{u}_t^N$. On the other hand, equation (8) presents a similar problem, as uncertainty about the level of the "natural" level of policy is not distinguishable from the level of the NAIRU. Putting it differently, there is no way policymakers can identify the NAIRU using the estimation of (7) and (8) without further assumptions (or, equivalently, the specification of further prior beliefs). Therefore, I assume that policymakers solve this identification problem by imposing the prior belief that on average, unemployment is equal to the NAIRU. That is to say that on average policy is at its "natural" level. This assumption provides a very natural way of estimating the NAIRU, which is to use univariate algorithms on the series of unemployment, in order to isolate the low frequency component. Furthermore, observe that this assumption is not contradictory and respects the coherence of the policymakers' model because, as will be shown in section 2.4 and appendix A, it correspond to a self-confirming equilibrium.

Conditional on their estimate of the NAIRU, policymakers can estimate the model's coefficients using standard regression methods. In this paper I consider three possible estimation algorithms: constant gain, ordinary and discounted least squares.

2.3.1 Constant gain (CG)

Following a large part of the most recent literature (see, among others, Sargent 1999, Williams 2003) in the baseline specification of the model I assume that policymakers update their beliefs using constant gain algorithms. These algorithms allow to update beliefs discounting the past and giving more weight to recent data. Recent data are considered more informative possibly because of the suspicion of drift in the parameters.

As mentioned above, the estimation works in two steps. In the first step policymakers obtain an estimate of the current value of the NAIRU using the following updating formula:

$$\hat{u}_{t|t}^N = \hat{u}_{t-1|t-1}^N + g_N R_{N,t-1}^{-1}\left(u_t - \hat{u}_{t-1|t-1}^N\right), \tag{10}$$

$$R_{N,t} = R_{N,t-1} + g_N\left(1 - R_{N,t-1}\right). \tag{11}$$

In the second step, policymakers use their estimate of the NAIRU to update their beliefs about the Phillips curve and aggregate demand coefficients:

[9] In equation (7), if average unemployment is equal to the non-inflationary rate, average inflation is equal to $\frac{\hat{c}_\pi}{1-\hat{\alpha}(1)}$ and there is no reason to assume it is equal to zero.

$$\hat{\beta}_t^i = \hat{\beta}_{t-1}^i + gR_{i,t-1}^{-1}x_t^i\left(y_t^i - x_t^{i\prime}\hat{\beta}_{t-1}^i\right), \tag{12}$$

$$R_{i,t} = R_{i,t-1} + g\left(x_t^i x_t^{i\prime} - R_{i,t-1}\right), \qquad i = \{\pi, u\}, \tag{13}$$

where $\hat{\beta}_t^\pi = [\hat{c}_\pi; \hat{\alpha}_1; \hat{\alpha}_2; \hat{\theta}_1; \hat{\theta}_2]'$; $y_t^\pi = \pi_t$; $x_t^\pi = [1; \pi_{t-1}; \pi_{t-2}; u_{t-1} - \hat{u}_{t|t}^N; u_{t-2} - \hat{u}_{t|t}^N]$; $\hat{\beta}_t^u = [\hat{c}_u; \hat{\rho}_1; \hat{\rho}_2]'$; $y_t^u = u_t - \hat{u}_{t|t}^N - V_{t-1}$; $x_t^u = [1; u_{t-1} - \hat{u}_{t|t}^N; u_{t-2} - \hat{u}_{t|t}^N]$. Notice from the expressions of the vectors x_t^π and x_t^u that policymakers approximate $\hat{u}_{t-1|t}^N$ and $\hat{u}_{t-2|t}^N$ with their last estimate of the current level of the NAIRU, $\hat{u}_{t|t}^N$.

g and g_N represent the constant gains. Observe that I allow for the possibility of different gain parameters in the NAIRU and coefficients' algorithms. The gain parameters control the rate at which new information affects beliefs. If g and g_N were decreasing and equal to $\frac{1}{t-1}$, equations (10), (11), (12) and (13) would be recursive representations of ordinary least squares estimates (if properly initialized).

2.3.2 Ordinary least squares (OLS)

A possible drawback of the constant gain algorithm is that policymakers do not get an estimate of the entire past series of the NAIRU, $\left\{\hat{u}_{s|t}^N\right\}_{s=1}^t$, to use in the second regression step. Therefore, here I propose an alternative two step algorithm. I assume that policymakers obtain an estimate of the series of the NAIRU using the following two-sided moving average of order $\bar{t}$ (which becomes one-sided at the beginning and the end of the sample):

$$\hat{u}_{s|t}^N = \frac{1}{\sum_{j=1}^t w_{j,t,s}} \sum_{j=1}^t w_{j,t,s} u_j, \qquad s = 1, ..., t$$

$$w_{j,t,s} = \begin{cases} 1, & \text{if} \quad |j - s| < \min\{t, \bar{t}\} \\ 0, & \text{otherwise.} \end{cases}$$

Conditional on these estimates of the NAIRU, the model's coefficients in (7) and (8) are estimated in the second step by running two simple OLS regressions.

2.3.3 Discounted least squares (DLS)

The first step of this algorithm is the same as the first step of the previous one. As in OLS, the advantage relative to CG is that the entire estimated series of the NAIRU is obtained in the first step and used in the second. Differently from before, I assume that the second step is estimated by the method of discounted least squares. What I call DLS is simply a weighted least squares method. The weight given to time s observation is equal to Δ^{t-s}, where t is the time period of the most recent data and $\Delta < 1$ is a discount factor. As CG, DLS gives more weight to recent data. Differently from CG, DLS does not have a recursive representation, which creates the disadvantage of additional computational time.

2.4 Equilibrium and Steady State

As is standard in most of the recent literature on learning, I focus on the concept of self-confirming equilibria.[10] In period t, policymakers form beliefs about the model's parameters. Their beliefs imply an optimal value for the policy variable V_t, which, in turn, affects the stochastic data generating process and, ultimately, next periods beliefs. This defines a map from today's beliefs to tomorrow's beliefs. A fixed point of this map is a self-confirming equilibrium. In other words, a self-confirming equilibrium is a situation in which policymakers' beliefs are expected not to change with the new vintage of data. Appendix A gives a formal definition and derives the model's self-confirming equilibrium for the baseline case in which beliefs are formed using the CG algorithm.

The key parameter that determines the existence of the self-confirming equilibrium is the unconditional variance of the NAIRU, i.e. $Var\left(u_t^N\right) = \frac{\sigma_\tau^2}{1-\gamma^2}$. When $\sigma_\tau^2 = 0$ it is easy to check that in the self-confirming equilibrium beliefs about the model's coefficients coincide with the true values. When $Var\left(u_t^N\right)$ is bigger than zero but "small", there still exists an equilibrium. In this case, equilibrium beliefs about the model's parameters do not necessarily coincide with the true model's parameters. The technical aspects of this result are discussed in appendix A. The intuition is the following: when $Var\left(u_t^N\right) > 0$ the mistakes associated with the estimation of the current level of the NAIRU[11] will bias toward zero the estimate the slope of the Phillips curve and of the persistence of unemployment deviations from the NAIRU in the aggregate demand equation. This affects also the other estimates, pushing them away from the true values. The distance between the true parameter values and the set of beliefs corresponding to the equilibrium is an increasing function of $Var\left(u_t^N\right)$. In other words, for small values of $Var\left(u_t^N\right)$ this distance is negligible, but becomes sizable for large values of $Var\left(u_t^N\right)$. When $Var\left(u_t^N\right)$ is very large with respect to σ_ε^2 and σ_η^2 the model does not admit a self-confirming equilibrium anymore.

I define the model's steady state as the unconditional mean of the stationary stochastic process for the vector $\left[\pi_t, u_t, V_t, u_t^N\right]$ in a self-confirming equilibrium. Observe that, in steady state, (5) implies $u_t = u^*$. Consequently, $V_t = 0$ follows from (2). Finally (9) implicitly defines the steady state inflation as a function of the equilibrium beliefs.

For simplicity, consider for a moment the intuitive case of a constant NAIRU.[12] Notice that, if $k = 1$, the steady state inflation is just the inflation target π^*. In other words, if policymakers do not wish to push unemployment below the NAIRU, the outcome is the optimal one in which unemployment is at its non-accelerating inflation level and inflation at its target. This corresponds to what Sargent (1999) calls the Ramsey outcome, i.e. the optimal time-inconsistent outcome of Kydland and Prescott (1977). If $0 \leq k < 1$ instead, the steady state inflation will be higher. This

[10] For a formal treatment of the issue, see Sargent (1999), Evans and Honkapohja (2001), Cho, Williams, and Sargent (2002), Williams (2003).

[11] There will always be these kind of mistakes, because policy is treating a stochastic process as a constant coefficient.

[12] As mentioned above and shown in appendix A, in this case the self-confirming equilibrium is simply given by the true parameters' values. Furthermore, for small values of σ_τ^2, the case of $\sigma_\tau^2 = 0$ provides a good approximation to the equlibrium beliefs.

corresponds to what Sargent (1999) defines as Nash inflation, corresponding to the time-consistent, non-optimal equilibrium of Kydland and Prescott (1977) and Barro and Gordon (1983).

To conclude, non-optimal equilibria can arise if policy targets an unemployment rate lower than the NAIRU, like in Kydland and Prescott (1977), Barro and Gordon (1983), Sargent (1999) and many others. These equilibria are characterized by an inflation rate higher than the optimal outcome. As mentioned in the introduction, the higher inflation property of the time consistent equilibria has been invoked by many authors in order to explain the high inflation episodes of recent US economic history. However, evaluating whether the time-consistency problem is quantitatively important in explaining low frequency movements of the inflation rate remains an open issue that I address in this paper. As it will be clear later, the data favor a model with a limited amount of inflation bias, induced by the time-consistency problem.

3 Interpreting the Great Inflation

In section 2.1 I have presented the baseline model of the paper and discussed the related technical issues. I now introduce a simplified version of the model, and use it to interpret and explain the postwar behavior of inflation and unemployment in the United States. Therefore, while the rest of the paper focuses on estimation and simulations, the objective of this section is to provide the intuition and the main ideas necessary to interpret the quantitative results.

3.1 A Special Case

As an illustrative example, consider the special case of the previous model given by the following simplified Phillips curve and aggregate demand equations:

$$\begin{aligned} \pi_t &= \pi_{t-1} - \theta(u_{t-1} - u^N) + \varepsilon_t && (14) \\ (u_t - u^N) &= V_t, && (15) \end{aligned}$$

where policymakers determine V by solving the following problem:

$$\begin{aligned} \min_{\{V_s\}} L &= E\sum_{s=t}^{\infty}\delta^{s-t}\left[\pi_t^2 + \left(u_t - k\hat{u}^N\right)^2\right], \\ \text{s.t.} \quad \pi_t &= \hat{c}_\pi + \hat{\alpha}\pi_{t-1} - \hat{\theta}(u_{t-1} - \hat{u}^N) + \widehat{\varepsilon}_t, && (16) \\ (u_t - \hat{u}^N) &= V_t. \end{aligned}$$

Observe that in order to obtain a closed form solution I have modified the timing of V in (15). The set of estimated parameters is given by $\hat{\beta} = [\hat{c}_\pi; \hat{\alpha}; \hat{\theta}; \hat{u}^N]$ and the states of the policy optimization problem are collected in $S_t = [1; \pi_t]$. The solution of the policy problem is given by the following linear control rule:

$$V_t = g(\hat{\beta})S_t = A(\hat{\beta}) + B(\hat{\beta})\pi_t, \quad (17)$$

where

$$A(\hat{\beta}) = -(1-k)\hat{u}^N + \frac{\left(1 + B(\hat{\beta})\frac{\hat{\alpha}}{\hat{\theta}}\right)\left(\hat{c}_\pi + \theta(1-k)\hat{u}^N\right)}{\hat{\theta}\left(1 + B(\hat{\beta})\frac{\hat{\alpha}}{\hat{\theta}}\right) - \left(\frac{\hat{\alpha}}{\hat{\theta}} - \frac{1}{\delta\hat{\theta}}\right)}$$

and

$$B(\hat{\beta}) = \frac{-\left(\frac{1}{\delta\hat{\theta}} + \hat{\theta} - \frac{\hat{\alpha}^2}{\hat{\theta}}\right) + \sqrt{\left(\frac{1}{\delta\hat{\theta}} + \hat{\theta} - \frac{\hat{\alpha}^2}{\hat{\theta}}\right)^2 + 4\hat{\alpha}^2}}{2\widehat{\alpha}}.$$

Observe that $B(\hat{\beta})$ is always positive (for positive values of $\hat{\alpha}$). This implies that policy reacts to inflation by pushing the unemployment rate upwards. Substituting (17) into (15) and the resulting equation into (14) we obtain the following equation for the inflation rate:

$$\pi_t = \theta\left(u^N - \hat{u}^N\right) - \theta A(\hat{\beta}) + \left(1 - \theta B(\hat{\beta})\right)\pi_{t-1} + \varepsilon_t. \tag{18}$$

This equation can be interpreted as a local approximation of inflation dynamics for given policymakers' beliefs about the state of the economy. Notice that the mismeasurement of the NAIRU shifts the mean of the inflation process. Instead, the strength of the policy reaction to inflation affects the persistence and therefore both the mean and the variance of the inflation process. In other words, the stronger policy reacts to inflation in the feedback rule, the more stationary and less volatile is inflation. In particular, it is easy to show that $B(\cdot)$ is a positive function of $\hat{\alpha}$ (the estimated persistence) for all positive values of $\hat{\alpha}$ and $\hat{\theta}$ (the estimated slope of the Phillips curve). This implies that the lower the estimated persistence of inflation in the Phillips curve, the lower the reaction to inflation and, given (18), the higher the actual inflation persistence.

Similar perverse dynamics can occur for wrong estimates of θ. In fact, $B(\cdot)$ is a positive function of $\hat{\theta}$ for values of $\hat{\theta} < \sqrt{\frac{1}{\delta} - \hat{\alpha}^2}$, which corresponds to a relevant set of possible values of $\hat{\theta}$. This means that the lower the estimated slope of the Phillips curve, the lower the reaction to inflation and, given (18), the higher the actual inflation persistence. That is, if policymakers perceive a costly inflation-unemployment trade-off, they will not be willing to accept higher unemployment for a limited relief from inflation. Therefore, they will react to inflation less strongly, ultimately leading to a less stationary inflation.

The extreme case of a unit root in the reduced form, univariate inflation process (18) occurs if policy does not respond to inflation at all. For example, this can happen if the perceived slope of the Phillips curve is zero. Figure 2 gives an idea of the shape of the function $B(\cdot)$. Figure 2a shows B as a function of $\hat{\alpha}$, when $\hat{\theta}$ is fixed to three different values. Besides the positive slope in all cases, notice the pronounced nonlinearity of B as a function of $\hat{\alpha}$, especially for high values of $\hat{\alpha}$. Figure 2b shows instead the opposite case in which $\hat{\alpha}$ is fixed to three different values and $\hat{\theta}$ is allowed to change. Interestingly, the effect of increasing $\hat{\theta}$ on B, heavily depends on the value of $\hat{\alpha}$: the higher the estimate of inflation persistence, the higher the effect of the estimate of the slope of the Phillips curve on the strength of the reaction to inflation.

It is important to realize that the case of weak policy reaction to inflation leads to inflation close to a unit root process. This is particularly dangerous for the stability of this economy. The reason is that if inflation is stationary, a mistake in the estimation of the NAIRU shifts the mean of the process. If inflation exhibits a unit root instead, a mistake in the estimate of the NAIRU creates a time trend in the inflation process.

It is easy to check that for this simple economy, the self-confirming equilibrium corresponds to $\hat{\beta} = \beta = \left[0; \alpha; \theta; u^N\right]$. It follows that the associated stationary stochastic process for the random vector $[\pi_t, u_t, V_t]$ is a simple VAR(1), whose coefficients are omitted for brevity. The steady state is $[\pi_t, u_t, V_t] = \left[\frac{(1-k)u^N}{\theta+B(\beta)+1-1/\delta}; u^N; 0\right]$. Notice that if $k = 1$ the steady state inflation is equal to zero. If $0 \leq k < 1$ instead, we have a positive inflation bias. For values of k close to one, the inflation bias is rather small.

3.2 A Simple Story for the Great Inflation

The Great Inflation refers to the high inflation and unemployment episode of the 60s, 70s and early 80s. At the end of the 50s and the beginning of the 60s inflation was lower than 2% (figure 1). However, in the early 60s inflation started increasing and reached its peak in 1981, at the level of more than 10%. After the beginning of the run-up of inflation, the unemployment rate (figure 1) remained low for more than ten years. Unemployment was lower than 6% until 1975 and between 1966 and 1970 unemployment was even lower than 4%. After 1975 the unemployment rate in the United States increased rapidly and remained high for more than ten years. Unemployment reached its peak at almost 11% in 1983. Summarizing, the data show a pronounced hump-shaped behavior for both inflation and unemployment, which strikes for dimension and duration. Moreover, as emphasized in the introduction, the disinflation was sharp and unemployment lags behind inflation during the entire duration of the Great Inflation episode.

The simplified model of the previous section provides a useful and powerful tool for the interpretation of this long and important episode of the postwar behavior of inflation and unemployment in the United States. In particular, the model helps to understand what was particular about that period of time. This is key in order to evaluate the possibility of future episodes of similarly bad economic performance.

I begin the imaginary simulation in 1960. Figure 3 plots real-time estimates of the NAIRU, inflation persistence in the Phillips curve and the slope of the Phillips curve, starting in 1960 and formed using data from 1948. These represent measures of real-time policymakers' beliefs affecting the choice of the policy variable V. Inflation persistence is measured by the sum of coefficients on lagged inflation, i.e. $\hat{\alpha}(1)$ in (7). The slope of the Phillips curve is measured as the sum of coefficients on unemployment deviations from the NAIRU, i.e. $-\hat{\theta}(1)$ in (7). Notice that the estimates of the NAIRU, inflation persistence and the slope of the Phillips curve are obtained using the specification of the policymakers' model of section 2 and not of the simplified model of the current section. Finally, these estimates are constructed using the baseline, constant gain learning algorithm of section 2.3.1, but their qualitative behavior is very robust to the alternative specifications of the learning algorithm.

3.2.1 The period of *overoptimism*

To start, notice that in the first part of the sample policymakers' estimates of the NAIRU were between 4% and 5%. These are low numbers, compared to our current estimates of the level of the NAIRU in the 60s and the first part of the 70s (see, for example, Orphanides and Williams 2002).

By now, it is not controversial that policymakers' real-time estimates of the NAIRU and potential output were too optimistic in the 60s and 70s. This fact is documented by many studies. For example, Orphanides (2000) focuses on the 70s and argues that the exceptionally high estimates of potential output were the main cause of the high inflation. DeLong (1997) and Romer and Romer (2002) describe in detail the fact that, already at the beginning of the 60s, "policymakers adopted a highly optimistic view of the levels of output and employment that could be reached without triggering inflation" (Romer and Romer 2002, p.20). According to the 1962 *Economic Report of the President* (EROP) a 4% rate of unemployment was considered not only a "reasonable and prudent" target, but also a "modest goal" for stabilization policy. The 1962 EROP also suggested the possibility that:

> stabilization policy alone could press unemployment significantly below 4 percent without creating substantial upward pressure on prices. (EROP 1962, p. 46.)

The 1963 EROP defined unemployment at 4% as an "unacceptable target". For several years the unemployment rate of 4% was used by the Council of Economic Advisers to define and construct an estimate of potential GNP. Still in 1967, the EROP stated that:

> The Council of Economic Advisers, among others, judged that an unemployment rate near 4 percent would ... yield approximate balance between the supply and demand for labor. (EROP 1967, p. 42.)

This erroneous belief that the NAIRU was so low led to overexpansionary monetary and fiscal policies. However, while this can explain why inflation started rising in the early 60s, it is not sufficient to explain why rational policymakers let inflation increase so much and for such a long period of time. What is key to rationalize policymakers' behavior in the 60s and 70s, is realizing that they were uncertain not only about the value of the NAIRU, but also about the value of all remaining parameters of their model. In particular, observe that the real-time estimate of inflation persistence in the early 60s was approximately 0.5 (figure 3b).[13] Remember that if the sum of coefficients on past inflation in the Phillips curve is less than one, this implies the existence of a long-run trade-off between inflation and unemployment.

The fact that in the 60s and early 70s most policymakers and economists believed in the existence of such a trade-off is not a mystery. DeLong (1997), Mayer (1999) and Romer and Romer (2002)

[13] The downward bias of this estimate might be due to several reasons, like the "craziness" of the 50s (due to the end of the World War II, the Korean war and the related decisions about price controls) or a structural break in the Phillips curve. A formal investigation of the causes is postponed until section 7.

stress the influence in the policymaking process of the original contribution of Samuelson and Solow (1960), which "presented the government with a menu of unemployment and inflation combinations among which it can choose" (Mayer 1999, p. 95). Even after the publication of Phelps' (1967) and Friedman's (1968) works on the natural rate hypothesis, the academic dispute about the existence of a long run trade-off continued for many years. For example, in 1968, at a symposium on inflation at New York University, three of the four distinguished speakers were in favor of the trade-off.[14] In that occasion, Robert Solow wrote:

> There remains, therefore, a genuine trade-off between inflation on one hand and employment and output on the other. ... For time spans that matter, there is no natural rate of unemployment. (Solow 1968, p. 14.)

Solow was not an exception. Blanchard and Fischer (1989) and Mayer (1999) note that, at least until the early 70s, most of the econometric studies estimated the sum of coefficients on past inflation to be less than unity. Therefore, given the academic climate, it is not surprising that policymakers tended to believe in the trade-off and in the low persistence of inflation in the Phillips curve. In 1971 the EROP wondered "Why is the inflation so stubborn" and showed surprise in arguing that:

> it is certainly also true that inflation has been and remains *exceptionally* [emphasis added] persistent. Observation of previous inflationary experience ... suggests that ... the rate of increase of prices is likely to decline after a moderate lag as slack in the economy emerges ... More sophisticated econometric analysis ... did not generally predict the rate of inflation experienced in 1970, given the actual conditions in 1970. (EROP 1971, p. 60.)

Summarizing, the high persistence of the inflation process and the absence of an exploitable inflation-unemployment trade-off were recognized by policymakers much too late. As shown above in the simplified version of the model, a low estimate of inflation persistence implies a low reaction to inflation, due to the wrong belief that inflation is rather stationary around some mean. Furthermore, if coupled with the mismeasurement of the NAIRU, this can create a time trend component in the inflation process, which closely resembles the actual path of inflation from 1963. I will refer to this period as the *overoptimism* period, during which inflation was perceived as stationary, the NAIRU as low and the optimal policy was keeping unemployment close to the estimated NAIRU, without much concern for the accelerating inflation.

3.2.2 The period of *overpessimism*

Slowly something changed. In fact, the model predicts that if policy does not react to inflation, the true stochastic process of inflation will be close to a unit root process, as equation (18) shows. On

[14] The three speakers in favor of the trade-off were Robert Solow, Albert Hart and James Tobin. The other speaker was Phillip Cagan.

the other hand, in the simplified model, if unemployment is kept close to the perceived NAIRU, estimating the Phillips curve is almost the same as estimating (18). Therefore, the estimates of inflation persistence will be slowly revised upwards, since (18) is close to a unit root process during the overoptimism period. As shown in figure 3b, the data strongly support the model's predictions.

However, while this would imply a reinforcement of the policy reaction to inflation, policy reaction remained low because of the following perverse mechanism: policymakers noticed that pushing unemployment above the *underestimated* NAIRU did not provide any relief from inflation. As a consequence they revised toward zero their beliefs about the slope of the Phillips curve ($\hat{\theta}$ decreases in the early 70s, as shown in figure 3c).[15] As implied by the model, this reduced the strength of policy reaction to inflation. In other words, even after the overoptimism period, policymakers kept reacting weakly to inflation, this time because they perceived a very costly inflation-unemployment trade-off. Clear evidence that this was actually the case is provided in the following statement by the EROP:

> When inflation failed to respond significantly to macroeconomic policy, a 90-day wage and price freeze was announced on August 15, 1971; it was followed by a period of mandatory wage and price controls. (EROP 1979, pp. 54-55.)

Commenting on the current economic conditions, the 1972 EROP further referred to a:

> tendency to an unsatisfactorily high rate of inflation which persists over a long period of time and is impervious to variations in the rate of unemployment, so that the tendency cannot be eradicated by any feasible acceptance of unemployment. (EROP 1972, p. 113.)

Even in 1979, the EROP wrote:

> We will not try to wring inflation out of our economic system by pursuing policies designed to bring about a recession. That course of action ... would be ineffective. Twice in the past decade inflation has accelerated and a recession has followed, but each recession brought only limited relief from inflation. (EROP 1979, p. 7.)

I will refer to this period as the *overpessimism* period, during which policy did not fight inflation because policymakers "did not believe it would work at an acceptable cost" (DeLong 1997, p. 264).[16] The *overpessimism* period is successive to the *overoptimism* one and accounts for the long duration of the hump-shaped episode. Notice that during this period, the model predicts that unemployment should remain rather low, increasing only partially in response to accelerating inflation.

[15] The technical reason is that standard estimation algorithms (like OLS), especially with a few available data, tend to move the constant and the slope of a regression equation in order to minimize the in-sample forecast error. When a break occurs in the constant of a regression equation, initially this will be attributed only in part to a shift in the constant and for the rest to a change in the slope.

[16] Recently, Cogley and Sargent (2003a) have reached a similar conclusion.

3.2.3 The disinflation

In the meantime, policymakers' estimate of inflation persistence in the Phillips curve had been updated toward the true value of one. Therefore, the episode ended after a proper revision of the estimate of the slope of the Phillips curve. This happened because of a sequence of new exogenous shocks, which caused updates of $\hat{\theta}$ toward the self-confirming equilibrium. When the bias of $\hat{\theta}$ decreased and the perceived inflation-unemployment trade-off improved, policymakers reacted strongly to high inflation, because they finally had a model of the economy that was approximately correct. Consequently, unemployment was pushed quickly way above the estimated NAIRU. The sharp disinflation was the result of this prompt and strong action. Policy maintained a high unemployment rate until inflation came back under control. At that point unemployment slowly returned to levels close to the NAIRU.

Notice that the model's predictions match very well the stylized facts. In fact, not only is the model able to account for the dimension and the duration of the episode, but it is also able to explain why the disinflation period was shorter than the run-up period and why unemployment increased and decreased during the 70s and early 80s, lagging behind inflation.

It is important to stress that this paper offers an explanation of the Volcker disinflation, which is not based on a sudden change of policymakers' preferences in the late 70s. Here, instead, the disinflation occurs when the inflation-unemployment trade-off becomes favorable, relative to the level of inflation.

4 Empirical Evidence

This section presents empirical evidence supporting the model of section 2 and the dynamics illustrated in section 3.

4.1 Statistical Evidence

In this section I focus on statistical and quantitative evidence to support the theory and the model presented above. I estimate the model of section 2 by maximum likelihood methods, over a sample period running from 1960:I to 2002:IV, using quarterly data on inflation and unemployment.[17] To compute the likelihood function I let policymakers estimate the approximate model and, consequently, choose V in every period. Putting the model in state space form, the likelihood can be computed using the Kalman filter (appendix B). The likelihood is maximized with respect to nine parameters, $\left[\alpha_1;\theta_1;\theta_2;\rho_1;\rho_2;k;\phi;\sigma_\varepsilon^2;\sigma_\eta^2\right]$, while I fix $\delta = 0.99$, $\pi^* = 2$, $\lambda = 1$, $u^* = 6$, $\gamma = 0.99$, $\sigma_\tau^2 = 0.0199$. Fixing δ is standard practise. I fix $\pi^* = 2$ because this number is close to the estimates of the inflation target level for the post-Volcker era (see, for instance, Bullard and Eusepi 2003, Schorfheide 2003, Favero and Rovelli 2003). However, differently from these previous

[17] As in the rest of the paper, inflation is measured by the annualized quarterly growth rate of the GDP deflator and unemployment by quarterly averages of the monthly civilian unemployment rate.

studies, there is no exogenous switch in the level of the inflation target in my model. In fact, I provide an explanation of the high inflation and disinflation of the 70s and early 80s, which is consistent with a policy inflation target of 2% also in the pre-Volcker period. λ is set to 1 to be consistent with some previous studies (for example, Sims 1988 and Sargent 1999).[18] Finally, for the coefficients of the exogenous process driving the NAIRU, $u^* = 6$ is chosen to match the average of unemployment during the sample period. γ is fixed at 0.99 and σ_τ^2 at 0.0199 to impose the prior belief that the NAIRU is smoother than unemployment itself. Notice that the value of γ and σ_τ^2 imply an unconditional variance of one for the NAIRU. Furthermore, in the case of the constant gain learning, g is set to 0.015 and g_N to 0.03. Observe that the constant gain for the estimation of the NAIRU is higher than the one for the estimation of the parameters. This captures the fact that policymakers expect the NAIRU to drift more than the model's coefficients. The constant gain algorithm is initialized by the estimates of the parameters and variances obtained in 1960 with DLS, using data from 1948. Finally, for the case of discounted least square learning, I choose to discount past data at the rate $\Delta = 1 - 1/120$. It is worth mentioning that the results of the paper are robust to these choices.

4.1.1 Estimation results

The point estimates and the standard errors of the free coefficients are reported in table 1 for the three specifications of the learning process. Four results stand out. First, the estimates obtained using different assumptions for the policymakers' learning process are very similar to each other. Second, the estimate of ϕ, the smoothing coefficient in the policy loss function, might appear as too high. However, once movements in the policy variable V are interpreted in terms of deviations of unemployment from the NAIRU, the size of ϕ is not a puzzle anymore.[19] Third, all the other coefficients have reasonable sign and size. Fourth, the estimate of k, the parameter that affects the unemployment target in the loss function (6), is close to unity, casting doubt on the inflationary bias story that I mentioned in the introduction.

Figure 4 plots the smoothed estimate of the NAIRU, which resembles previous estimates in the literature (Staiger, Stock, and Watson, 1997 and 2001, Gordon, 1997). The only notable difference is the fact that the NAIRU is estimated to be slightly higher than standard estimates in the period between 1965 and 1975. The reason for this difference is the fact that this paper does not use only a Phillips curve to estimate the NAIRU, but also an aggregate demand relation. Conditional on the model being true, this should lead to better estimates.

Figure 5 plots the evolution over time of the model's policy variable V. A measure of the real

[18] Interestingly, if λ is treated as a free parameter in the estimation procedure, the point estimate is 0.99, with a standard error of 0.36.

[19] For example, consider a permanent unitary increase in the level of V_t. This would create a permanent deviation of unemployment from the non-inflationary rate equal to $\frac{1}{1-\rho(1)}$. Therefore the term ϕV_t^2 in the loss function could be expressed in terms of deviations of unemployment from the target as $\phi\left(1-\rho(1)\right)^2\left(u_t - \hat{u}_t^N\right)^2$, where the weight becomes $\phi\left(1-\rho(1)\right)^2$. If unemployment is very persistent, $\rho(1)$ will be close to one and $\phi\left(1-\rho(1)\right)^2$ will be a small number, despite a very high ϕ.

rate of interest (rescaled)[20] is reported for comparison. Notice that, especially in the second part of the sample, the two series are very similar. This is remarkable, if we consider that the time series for V has been obtained without any information related to the interest rate. However, sizable differences among the two series are evident in the first part of the sample, suggesting that focusing only on the evolution of the real interest rate as an indicator of policy might be restrictive.

4.1.2 Model's fit

In order to evaluate the fit of the model, I compare it to the fit of unrestricted multivariate linear models, i.e. bivariate vector autoregressions (VAR) with inflation and unemployment, similar to the ones used in King and Watson (1994). I consider the three alternatives of a VAR(2), a VAR(3) and a VAR(4), where the number in parenthesis indicates the number of lags included in the VAR. Table 2 reports three measures of fit for the learning models and the reference ones. The first column of table 2 reports the value of the log-likelihood, evaluated at the peak. The log-likelihood is useful, but certainly not very informative in a case in which the candidate models have such a different number of free parameters. A VAR(2) has 13 free parameters, a VAR(3) has 17 and a VAR(4) has 21, as opposed to the 9 free parameters of my model. The difference in the number of free parameters is so large that, even though the models are non-nested, it is reasonable to expect a lower log-likelihood for the learning models with respect to the reference ones.

To solve this problem, the second column of table 2 reports the Bayesian Information Criterion (BIC). BIC is an asymptotic approximation of the marginal likelihood, which, in turn, is proportional to the posterior probability of the model, under flat prior on the models' space. BIC automatically penalizes models with higher number of parameters.

Column 3 of table 2 is in principle the most reliable measure of fit. The last column in fact reports the logarithm of the marginal likelihood itself, computed using the Laplace method, based on a second order approximation of the posterior around the peak. For the VAR models instead, it is possible to compute the exact marginal likelihood. Appendix C illustrates these points. The potential drawback of the marginal likelihood is that it requires the specification of proper (and therefore, potentially informative) priors for the models' parameters. The reason is that the marginal likelihood is basically the integral of the posterior of the model's parameters. Therefore, the scale of this posterior (and consequently of the prior) matters in comparing models.

To remain as agnostic as possible, for the model comparison I use training sample priors. The way training sample priors work is by extending the estimation sample (usually backwards). The likelihood of the training sample is then interpreted as a prior density for the original sample. If the size of the training sample is large enough, the prior generated in this way will be proper. At this point, the key step of the procedure is scaling the prior in order to integrate to one.[21] Therefore,

[20] The real rate of interest is computed as federal funds rate minus the quarterly inflation rate averaged over the last four quarters.

[21] Computing the integrating constant of the prior is easy in the case of the VAR models, since we know the exact formula. For the case of the learning models instead I use the Laplace approximation formula, since we do not have an analytic expression for the prior density (see appendix C for details).

while the original sample starts in 1960, I extend it back to 1953 in order to include a training sample. For the purposes of this section it is enough to observe that the posterior mode of the models' parameters does not show important differences from the maximum likelihood estimates, reported in table 1.

From table 2 it is clear that, as expected, the log-likelihood of the learning models is lower than the unrestricted alternatives. However, once the different number of parameters is taken into account, it seems that the learning models fit the data better than the reference VARs. In particular, the marginal likelihood favors the learning models over the alternative ones. Observe that the results of the model comparison exercise are robust to different policymakers' learning schemes.

4.2 Simulations

This subsection considers quantitative simulations of the model in the case of the benchmark (CG) specification. The purpose of these simulations is twofold. First, they show that the model produces a pronounced, prolonged and asymmetric hump-shaped behavior in inflation and unemployment. In addition, these simulations highlight that the mismeasurement of the NAIRU or a too low unemployment target are not sufficient to reproduce this kind of hump-shaped behavior.

In order to do so, I conduct the following exercise. I start in 1960:I, assuming that data on inflation and unemployment are the actual data from 1948:I to 1959:IV. In 1960:I, policymakers optimize their objective function, on the base of the current estimates and fix the value of V for the current period. Unemployment and inflation in 1960:II are determined through the true Phillips curve and aggregate demand equations, (5) and (2). In the next period, policymakers reestimate their approximate model and choose the new value for the policy variable. With this mechanism I simulate 32 years of quarterly data, up to 2002. In the simulations I assume that the true value of the model's parameters are the values estimated in the previous subsection and reported in the first column of table 1. Also as "true" NAIRU I use the smoothed estimate plotted in figure 4. Finally, I perform the simulations generating sequences of *i.i.d.* random innovations with mean zero and variance corresponding to the estimated value reported in table 1. Even if it is assumed that policymakers update their beliefs using the constant gain algorithm of section 2.3.1, the qualitative results are robust to the choice of alternative policymakers' learning schemes.

I perform $3,000$ simulations using identical initial conditions in 1960, but different series of exogenous shocks. The main results are summarized in figure 6. Figures 6a-d graph the empirical distribution of four objects: the maximum level reached by the inflation rate between 1960 and 2003 ($\max(\pi)$); the time period in which inflation reaches its maximum level (t_π^*); the maximum level reached by the unemployment rate between 1960 and 2003 ($\max(u)$); the difference (expressed in years) between the time period in which unemployment peaks and the time period in which inflation peaks ($t_u^* - t_\pi^*$). The scatter plots of figures 6e and 6f are meant to illustrate two bivariate relations: the relation between peak time and the peak level of inflation; the relation between the peak level of unemployment and inflation. If we neglect 1974 (which is contaminated by the effect of the first oil price shock), actual inflation peaked in 1981 at a level 10.55%. Actual unemployment peaked two years later, at a level of 10.67%. Overall, figure 6 makes clear that the model reproduces remarkably well the main characteristics of the Great Inflation. In particular, the model fully captures the dimension, the duration of the high inflation episode and the fact that unemployment peaks after inflation.

Showing that, on average, the simulated time paths exhibit a rapid disinflation is less straightforward, because simply averaging the simulated time paths would not preserve the typical shape. To solve this problem I compute the average path, but only after rescaling the horizontal and vertical axis in every simulation, to make them peak at the same time and at the same level.[22] The result is plotted in figure 7, where the peak time of the average inflation path is normalized to zero and the peak value to the peak level of actual inflation in 1981. Actual inflation is also reported for comparison. The striking feature of figure 7 is that the average of the simulated paths captures perfectly the so called Volcker disinflation, that many macro models have difficulties in addressing.

Figure 8a plots a simulation of the behavior of inflation and unemployment, when the standard deviation of the shocks to the Phillips curve and the aggregate demand are chosen as small as possible and are fixed respectively to 0 and 0.005.[23] This is clearly improper, since the model is nonlinear and the size of the shocks affects the speed of the learning process. Nevertheless, if we keep this in mind, this exercise is useful because it shows that the model can capture the low frequency behavior of inflation and unemployment, even with small exogenous shocks (and stress further the role of initial conditions, which start these peculiar dynamics). A few results stand out. First, notice the large and long hump-shaped behavior of inflation, which increases from about 1% in 1961 to a level of about 11% between 1980 and 1985. The rate of increase is approximately constant. Second, unemployment lags behind inflation, peaking a few periods after the inflation peak. Third, the disinflation is rapid and starts when unemployment begins to increase significantly. Unemployment starts increasing fast after the revision of the estimate of the slope of the Phillips curve. Figure 9 plots policymakers' real-time estimates of the NAIRU, inflation persistence and the slope of the Phillips curve, for these simulation exercise. The similarities between figure 9 and figure 3 (showing the actual real-time estimates) are remarkable, even though the estimate of the slope of the Phillips curve seems more sensitive to the actual sequence of exogenous shocks than the estimates of the NAIRU and the inflation persistence.

To highlight the important determinants of such behavior, figures 8b and 8c plot the results of counterfactual simulation exercises. Here I use the same sequence of random shocks used to generate

[22] In order to determine the peak time of inflation in a robust way, I compute a five years moving average for every simulated path and select the point in time in which the resulting smooth series reaches the maximum.

[23] Two observations are necessary at this point. First, the standard deviation of the shocks to the aggregate demand cannot be set to zero because some random variation in the regressors of the Phillips curve is necessary for a meaningful learning dynamics and the convergence to the self-confirming equilibrium. Second, the simulations with low or zero variance shocks look all exactly the same except for the time period in which inflation (and, of course, unemployment) peaks.

figure 8a, but I change some of the values of the parameters of the model. In the simulation of figure 8b I assume that policymakers know the exact value of the slope of the Phillips curve. However they are uncertain about the NAIRU and inflation persistence in the Phillips curve and they have to estimate them. It is evident that the high inflation episode would have not disappeared, but would have lasted less long. In fact, in this case inflation peaks much earlier at a level of 6%. It is important to notice that, consistently with the predictions of the model, in figure 8b the rapid disinflation disappears. In figure 8c instead, I assume that policymakers have the correct estimate of the parameters of the Phillips curve, except for the NAIRU, which is estimated in the usual way. It is clear that, if this had been the case, inflation would have increased much less than it did. Indeed, this graph does not exhibit any sizable low frequency behavior. From this counterfactual exercises, I conclude that the mismeasurement of the NAIRU is a fundamental ingredient for the generation of the high inflation episode of the 60s and 70s. However, even if necessary, the mismeasurement of the NAIRU alone is not sufficient to explain the low frequency behavior of inflation and unemployment of the last forty years. In fact, optimizing policymakers with the correct parameters of the economy in hand would have successfully kept inflation under control.

Summarizing, these simulations show the importance of learning dynamics for the explanation of the high inflation and unemployment episode of the 60s, 70s and early 80s. A model with a realistic amount of inflation bias or/and a model in which policymakers estimate the NAIRU in real time are not able to generate hump-shaped episodes like the one observed in the data. In order to generate a quantitatively similar episode, the learning on the parameters of the Phillips curve seems essential.

As a further check of the importance of learning dynamics, I reestimate the model under the last assumption that policymakers know all the true coefficients of the Phillips curve except for the value of the NAIRU. The log-likelihood, the BIC and the log-marginal likelihood of this model are respectively equal to -248.59, -271.72 and -271.29. These values are slightly higher than the respective values correspondent to the baseline CG model, reported in the first row of table 2. This would seems to suggest that the model without learning generates an improvement in fit over the baseline model. However, there is one dimension along which the model without learning seems to completely fail. In fact, the smoothed estimate of the NAIRU obtained using this model (figure 10) totally differs both from any estimate previously obtained in the literature (see, for example, Gordon 1997, Staiger, Stock, and Watson 2001) and from the common intuition that the NAIRU increased in the 60s and 70s and decreased afterwards. In other words, taking explicitly into account any plausible prior beliefs about the behavior of the NAIRU over time would heavily penalize the model without learning.

4.3 The fit of alternative explanations

As shown in the previous section, the learning model is able to reproduce remarkably well the postwar behavior of inflation and unemployment in the United States. The purpose of this section is to compare the fit of the baseline learning model to the fit of some alternative explanations of the Great Inflation.

As I mentioned in the introduction, most of the alternative theories of the Great Inflation adopt an exogenous shift in the preferences of policymakers to explain the sharp disinflation and the low level of inflation since the early 80s. On the other hand, the great advantage of the approach taken in this paper is the fact that the time variation in the conduct of stabilization policy is fully endogenized. To address this issue and assess the importance of the exogenous and unexplainable channels of time variation in policy, I estimate a version of the learning model where also the parameters of the loss function are allowed to vary over time. In particular, in this subsection, I will assume the following form for the loss function of policymakers:

$$L = \hat{E}\sum_{s=t}^{\infty}\delta^{s-t}\left[(\pi_t - \pi_t^*)^2 + \lambda_t\left(u_t - k_t\hat{u}_t^N\right)^2 + \phi\left(V_t - V_{t-1}\right)^2\right],$$

which differs from (6) because three of the parameters representing the preferences of policymakers (k, π^* and λ) are allowed to change over time. Following many previous studies (for example, Clarida, Gali, and Gertler 2000), I model the time variation in k, π^* and λ in the simplest possible way, i.e. allowing them to differ between the pre and post-Volcker period:

$$[\pi_t^*, k_t, \lambda_t] = \begin{cases} [\pi_1^*, k_1, \lambda_1] & \text{for } t < \bar{t} \\ [\pi_2^*, k_2, \lambda_2] & \text{for } t \geq \bar{t}, \end{cases}$$

where $\bar{t}$ is set to the fourth quarter of 1979.

The estimates of the model with a break in the preferences of policymakers in addition to learning are reported in the first column of table 4, together with the value of the log-likelihood and the BIC.[24] Notice that the parameter controlling the size of the inflationary bias (k) does not change between the pre and post-Volcker periods. Instead, as expected, both the target for inflation and the weight on the unemployment objective seem to decrease after 1979, although the estimates are quite imprecise. In particular, the p-values for the tests with null hypothesis of no change in π^* and λ are respectively 0.5 and 0.34, implying a failure to reject the null of no change in both cases. This is confirmed by the relatively small improvement in the log-likelihood and the substantial deterioration of the BIC.

As a further check, I consider another alternative, which is a model without learning on the parameters of the Phillips Curve (except for the NAIRU), but with shifts in the policymakers' preferences. The estimates of this model are reported in the second column of table 4. Notice that in this case k is estimated to be zero, both in the pre and post-Volcker periods. On the other hand, both π^* and λ, seem to decrease in the second part of the sample. The changes in the values of π^* and λ before and after 1979 are even larger than in the previous case, but they are also even less precisely estimated. In fact, the p-values for the tests with null hypothesis of no change in π^*

[24]The marginal likelihood for this model is not computed, since a prior for the preference parameters of the post-Volcker period cannot be obtained using a training sample prior (and, therefore, the results would not be easily comparable to the last column of table 2).

and λ are respectively 0.52 and 0.5, implying again a failure to reject the null of no change in both cases. The log-likelihood and the BIC improve over the previous specification, but the BIC is still inferior with respect to the BIC of the baseline learning model.

The overall impression is that the fit of the baseline model is superior to the alternatives presented in this subsection. In other words, once policymakers are assumed to learn over time about the structural relations of the economy, the time variation in the policy preferences seems redundant.

5 Policymakers' Learning and Forward Looking Agents

In the model of section 2 I used the Phillips curve (1) that can be rewritten as the backward looking equation (5), which is similar to the one estimated by policymakers. In other words I have assumed that policymakers have the correct model of the economy in hand, but are uncertain about the value of the model's parameters. I have argued that this assumption simplifies the analysis and favors the intuition because it helps to isolate the role of policymakers' learning in the generation of the high inflation episode of recent US economic history. This section tests the robustness of the results to alternative popular assumptions about the price setting equation. In particular, here I will concentrate on the case in which price setters are forward looking (for example, as in Calvo 1983). This assumption generates a forward looking Phillips curve, in which today's prices and inflation depend on expectations of future prices and inflation. For symmetry and in the spirit of the recent new-Keynesian literature, I will also assume that the aggregate demand equation has potentially a forward looking component. To be concrete, in this subsection I will replace the Phillips curve and the aggregate demand, (5) and (2), with the following forward looking versions:

$$\pi_t = \frac{1}{1+d}\pi_{t-1} + \frac{d}{1+d}E_{t-1}\pi_{t+1} - \theta E_{t-1}\left(u_t - u_t^N\right) + \varepsilon_t, \tag{19}$$

$$u_t - u_t^N = \rho_b\left(u_{t-1} - u_{t-1}^N\right) + \rho_f E_{t-1}\left(u_{t+1} - u_{t+1}^N\right) + V_{t-1} + \eta_t, \tag{20}$$

where d is the private sector's discount factor. The first equation is a forward looking Phillips curve and, similarly to Christiano, Eichenbaum, and Evans (2001), can be derived from the firms' profits maximization problem. As in Calvo (1983), firms are assumed to be able to reoptimize and set their prices with a certain probability in every period. If they do not reoptimize, they are assumed to set their prices with an indexation mechanism to past inflation. This simple indexation rule in the price setters behavior explains the backward looking component in (19). As pointed out in many studies, this component helps to account for the high degree of inflation persistence observed in the data (see, for instance, Fuhrer and Moore 1995a, Gali and Gertler 1999, Christiano, Eichenbaum, and Evans 2001, Smets and Wouters 2003). Observe that (19) is a vertical Phillips curve in the long run, implying absence of long run trade-off between inflation and unemployment.

The second equation is a forward looking aggregate demand equation. This expression can be derived from the household maximization problem, allowing for external habit formation, as in Smets and Wouters (2003), when $\rho_b + \rho_f = 1$ and $\rho_b < 0.5$. Habit formation explains the backward looking component, which helps to account for the persistence observed in the data. In my estimation and simulations I will relax the previous equality and inequality restrictions on ρ_b and ρ_f because those restrictions seem to be at odds with the data. Observe that the value of V is chosen by the policymakers in every period in the same way of section 2. In other words, policymakers estimate the backward looking model given by (7) and (8) and fix V solving the linear quadratic problem based on their current beliefs. Differently from the benchmark case, policymakers not only are using estimated parameters instead of true ones, but they are also using a model structure which is fundamentally different from the true model of the economy.

In order to solve the model and possibly estimate it, we need to make assumptions on the private sector's expectation formation process. As in Sargent (1999), I assume that the private sector knows exactly the way policy is made. In other words, people know that policymakers estimate (7) and (8) and fix V by solving the linear quadratic problem (6). However, there are many possible assumptions for the way agents form their expectations on future policy. For completeness, I analyze two alternative (and completely opposite) cases. As a starting point, following most of the adaptive learning literature, I assume that the private sector thinks that policymakers will not revise their estimates of (7) and (8) in the future and will continue to implement policy based on their latest estimates of (7) and (8). For simplicity, I will denote this behavior of the private sector as "partially rational". As an alternative and complication of the baseline hypothesis, I also analyze the case in which the private sector is "fully rational" and takes into account the fact that policymakers will revise their estimates of (7) and (8) on the base of future data.

5.1 Partially Rational Private Agents

The solution of this first specification of the model is reasonably standard. In fact, under the assumptions that the coefficients $g(\hat{\beta})$ in (9) remain constant in the infinite future, equations (19), (20), (9) and (3) form a *linear* rational expectations system of equations (RESE), which can be solved using standard methods (like King and Watson 1998, Sims 2001a). In reality the coefficients $g(\hat{\beta})$ change, therefore the RESE must be solved in every period of the sample to find the time varying data generating process. Given a random sequence of exogenous, non-policy shocks, this method can be used to construct simulated paths of the variables of interest. Given the actual data, this method can be used to compute the likelihood and maximize it with respect to the model's parameters. I fix the private sector discount factor (d) to 0.99 and maximize the likelihood with respect to seven parameters $\left[\theta; \rho_b; \rho_f; k; \phi; \sigma_\varepsilon^2; \sigma_\eta^2\right]$.

The log-likelihood and BIC for the forward looking model are reported in table 5. Even though the data seem to favor the backward looking specification, it is worth focusing on the estimates of the forward looking model. The point estimates and the standard errors of the parameters are reported in table 5.[25] First notice that the point estimate of θ has the correct sign, but it is very small.

[25] As in the benchmark case, here I assumed that policymakers learn using a constant gain algorithm.

However this must be expected. The reason is that, if the private sector thinks that the policy rule parameters will not change in the future, a higher θ would imply an immediate enormous effect on inflation's expectations and, consequently on inflation itself. We do not observe this in the data. The second thing to notice is that the weight on the forward looking component of the aggregate demand equation is estimated to be zero. The statistical evidence for this result is very strong. In fact, when I reestimate the model imposing the restriction $\rho_f > 0.5$, the log-likelihood drops dramatically. This makes equation (20) hard to reconcile with the household optimization behavior and supports the skepticism toward the empirical importance of forward looking consumers' behaviors shown, among others, by Estrella and Fuhrer (2002) and Fuhrer and Rudebusch (2002).

As in the previous section, I fix the parameters to their estimated value and perform $3,000$ simulations of the pattern of inflation and unemployment by generating random sequences of exogenous shocks. The second column of table 3 reports medians, lower and upper quartiles of $\max(\pi)$, $\max(u)$, t_π^* and $t_u^* - t_\pi^*$. The median peak level of inflation is even higher than for the baseline model, unemployment seems to peak a bit later and there is more uncertainty related to the peak time of inflation. However, overall, the summary statistics are in line with the actual behavior of inflation and unemployment during the Great Inflation. The asymmetry of the high inflation episodes generated by the forward looking model is demonstrated in figures 11a and 11b, which are constructed in the same way as figures 8a and 7.

Summarizing, the results obtained with the baseline model are very robust to the use of the forward looking Phillips curve. We observe both inflation increasing and decreasing faster, as well as unemployment peaking after inflation.

5.2 Fully Rational Private Agents

The solution of the model with the full rationality assumption is non-standard. The reason is that the system given by (19), (20), (3), (9), (10), (11), (12) and (13) is a *nonlinear* RESE, because the coefficients $g(\hat{\beta})$ are nonlinear functions of the endogenous variables. To solve the model I use numerical methods. In particular I apply the method that Fackler and Miranda (2001) propose for a general nonlinear RESE. It consists in approximating unknown functions (which, in our case, are the expectations $E_{t-1}\pi_{t+1}$ and $E_{t-1}u_{t+1}$) with projection techniques. In other words, the response function is approximated by the function Φs_{t-1}, where Φ is a matrix of coefficients and s_{t-1} is a collection of basis functions of the state variables. The method consists of replacing $E_{t-1}\pi_{t+1}$ and $E_{t-1}u_{t+1}$ with the approximations and finding the matrix of coefficients Φ^* that solve the RESE.[26] The difficulty here is the high dimension of the problem and of the state vector. The details of the solution method are given in Appendix D. As far as I know, this is the first attempt to solve numerically a model with forward looking agents and policymakers' learning.

[26] Fernandez-Villaverde and Rubio (2002) propose a similar method for the solution of nonlinear RESE. However, they use a piece-wise linear approximation of the solution, which in this case is unfeasible for the high number of potential states of the problem. Older techniques are the ones presented in Fair and Taylor (1983 and 1990). However their method is particularly suitable for perfect foresight solutions, which are clearly undesirable for the solution of a model with learning.

The solution of the model can take hours, therefore estimation, though theoretically easy, is not practical.[27] For this reason I perform only a set of simulations, calibrating the model in the following way: as before, d is fixed to 0.99; the parameters $\left[\rho_b; \rho_f; k; \phi; \sigma_\varepsilon^2; \sigma_\eta^2\right]$ are fixed to the point estimates in the model's specification of section 5.1; θ is fixed to 0.01, which seems a reasonable value. Notice that θ is fixed to be higher than the point estimate in the previous specification. The reason is that, under this new assumption on the expectation formation, the private sector recognizes that policy mistakes will be slowly corrected in the future. Consequently, observing policy mistakes today does not cause a huge effect on expected inflation and inflation itself.

I perform $3,000$ simulations and report medians, lower and upper quartiles of $\max(\pi)$, $\max(u)$, t_π^* and $t_u^* - t_\pi^*$ in the last column of table 3. Observe that inflation and unemployment continue to peak at high levels, which are comparable to what we observe in the data. However, on average, inflation peaks earlier and difference between the peak time of unemployment and inflation is higher than in the simulations with the other models. This is confirmed by figure 11c and 11d.

Generally, two points stand out from the analysis of table 3 and figures 11c and 11d: first, the simulations of this model continue to exhibit the pronounced and prolonged hump-shape behavior of inflation and unemployment; second, the typical simulated path of inflation peaks earlier and does not show a disinflation as sharp as the one observed in the data and reproduced by the baseline model and the forward looking model of section 5.1. However, this is not at all surprising, since the private sector is assumed to be forward looking and to take fully into account future policy changes. If the private sector knows the way policy is made, it will predict that policy will become very active against inflation as soon as policymakers' estimates of the Phillips curve will be revised. Inflation expectations will be automatically adjusted to take this into account. Consequently, inflation will reflect these adjustments in expected inflation.

This section suggests two conclusions. First, that even in the forward looking Phillips curve case, introducing policymakers' learning is important to explain the dimension and duration of high inflation episodes. Second, that the assumption of fully rational agents, who form their expectations taking into account the future evolution of policymakers' beliefs about the economy, is probably too strong and at odds with the data on the disinflation episode of 1981.

6 Extensions

In this section I analyze two additional extensions of the benchmark framework presented in section 2. Recognizing that the heteroskedasticity is an important characteristic of the exogenous shocks to the economy in the last forty years, the first subsection introduces stochastic volatility in the baseline model of section 2. The second subsection instead analyzes the case of sophisticated policymakers that estimate the NAIRU using all their information more efficiently. These are important extensions and robustness checks, but have been so far postponed because they are

[27] Since it would require several likelihood evaluation and, therefore, several model's solutions for different values of the parameters.

technically much more demanding with respect to the benchmark model.

6.1 Heteroskedasticity and the Role of Non-Policy Shocks

The fact that exogenous shocks have exhibited high heteroskedasticity over the last forty years has been highlighted in many studies (see, for instance, Blanchard and Simon 2001, Cogley and Sargent 2003b, Kim and Nelson 1999a, Kim, Nelson, and Piger 2001, McConnell and Perez-Quiros 2000, Primiceri 2003, Sims and Zha 2004, Stock and Watson 2002, Stock and Watson 2003b). Furthermore, many authors have argued that one of the main causes of the American inflation in the 70s was exactly a particularly bad sequence of exogenous, non-policy shocks (see, for instance, Blinder 1979). On the other hand, this paper has so far provided an explanation of the high inflation episode based on the policy behavior. Therefore, it seems natural to extend and reestimate the model of section 2 allowing for heteroskedastic innovations. In this way, the model leaves it up to the data to determine whether the channel of the policymakers' behavior remains important once the possibility of time varying variances is taken into account.

There is also another reason why it is potentially very important to take heteroskedasticity into account. This is because, in a learning model, the size of the exogenous shocks affects the speed of the learning process. In very general terms, higher volatility of an equation's disturbances would slow down the learning process. On the other hand, higher volatility of the regressors would speed up the learning process.

In order to allow for heteroskedasticity, I make the assumption that the standard deviations of the exogenous innovations of equations (1) and (2) follow a geometric random walk process:

$$\log \sigma_{\varepsilon,t} = \log \sigma_{\varepsilon,t-1} + \nu_{\varepsilon,t}, \tag{21}$$

$$\log \sigma_{\eta,t} = \log \sigma_{\eta,t-1} + \nu_{\eta,t}. \tag{22}$$

This class of models is known as stochastic volatility models and constitutes an alternative to ARCH and GARCH models. The crucial difference is that the variances generated by (21) and (22) are unobservable components.

The estimation of the model augmented with stochastic volatility is considerably more involved than the benchmark case. I adopt Markov chain Monte Carlo (MCMC) methods for the posterior numerical evaluation of the parameters of interest. MCMC deals efficiently with the nonlinearities of the model, allowing to draw from lower dimensional and standard distributions as opposed to the high dimensional joint posterior of the whole parameters set. Appendix E gives the details of the estimation algorithm.

The fourth column of table 1 reports the posterior median and the posterior standard deviations of the parameters of interest. They are similar to the benchmark case of the previous section. Figure 12 plots the posterior median (and the 68% error band) of the time varying standard deviations of the innovations to the Phillips curve and the aggregate demand equation. Observe that the time path is consistent with the past literature (see, for instance, Cogley and Sargent 2003b, Primiceri

2003, Sims and Zha 2004, Stock and Watson 2002). The variance of the innovations to the Phillips curve is low in the 60s, high in the 70s and early 80s and low again since 1985 to the end of the sample. The variance of the demand shocks in the aggregate demand equation follows approximately the same time pattern. Notice the sharp decrease in the standard deviation of the shocks to the unemployment rate that occurred in the early 80s, a phenomenon which is known as the Great Moderation.

I evaluate the role of the exogenous shocks in the high inflation episode of the 70s by performing counterfactual simulation exercises. The methodology I adopt is straightforward. I fix the model's coefficients to the estimated posterior medians and reconstruct an estimate of the sequence of exogenous shocks $\{\varepsilon_t\}_{t=1}^T$ and $\{\eta_t\}_{t=1}^T$. Starting from 1960:I, these shocks can be used to simulate counterfactual data, constructed using different values of the parameters. These new series can be interpreted as the realization of the data that would have been observed, had the parameters of the model been the ones used to generate the series. In this context, the interesting experiment consists of replaying history assuming that the standard deviations of the exogenous shocks were lower that the estimated standard deviations of the 60s and 70s. For comparison, figure 13a plots actual inflation and unemployment. Figure 13b plots the counterfactual paths of inflation and unemployment when the time varying standard deviations of supply and demand shocks are both replaced by their value in 1995, which is one of the less volatile periods. It is clear that the differences from the actual behavior of inflation and unemployment are minor, except, of course, for the reduced volatility. Figure 13c and 13d plot the counterfactual paths of inflation and unemployment when the time varying standard deviations of respectively supply and demand shocks are replaced by their value in 1995. Figure 13d shows that the worst scenario for inflation would have occurred in the case of reduced volatility of the shocks to unemployment. The reason is that it would have slowed down the updating process toward the self-confirming value for the beliefs about the slope of the Phillips curve parameter. For the same reason, the best scenario would have been the one of figure 13c, i.e. the case of reduced volatility of the supply shocks and the estimated historical volatility of the demand shocks. Finally notice that, although the behavior of inflation differs across the four plots, even in the most favorable scenario of figure 13c, inflation would still have peaked at the high level of about 9%.

Overall, this section suggests two conclusions. First, allowing for heteroskedasticity of the exogenous shocks does not undermine the results that the low frequency behavior of inflation and unemployment is largely explained by the evolution of policymakers' beliefs. Second, in the context of this model, allowing for time varying variances does not generate important differences in the speed of the policymakers' learning process and the timing of the disinflation. Instead, the dynamics seem to be mainly driven by the convergence process of policymakers' beliefs to the self-confirming equilibrium, starting from the initial conditions. This finding also facilitates the interpretation of figures 8a, 11a and 11c.

6.2 Prior Beliefs and Optimal Use of Information

So far I have analyzed and estimated a model in which policymakers' estimate the NAIRU by univariate algorithms. I have argued that this assumption simplifies the analysis importantly. Moreover, it seems a very realistic assumption from an historical perspective. Nevertheless, it is useful to answer the following question: would a less naive behavior of policymakers have prevented the rise and fall in inflation and unemployment. In other words, would better policymakers have generated a more satisfactory outcome? The objective of this subsection is to answer this question. Looking ahead, this subsection demonstrates that, while possible, it is very unlikely that policymakers behaving in a more sophisticated way would have led to a very different outcome.

To do so, in this subsection I drop the univariate algorithms assumption for the estimation of the NAIRU. Instead, policymakers are assumed to be fully rational and to use all the available information in a model consistent way. In particular, policymakers estimate equations (7) and (8) in every period by maximum likelihood. They specify a well defined exogenous stochastic process for the NAIRU and evaluate the likelihood by the Kalman filter. Since policymakers use their current estimates as true values in the optimization problem, it would not make sense to assume that they know the true stochastic process (3) for the NAIRU, which is mean reverting. Instead, I assume that they estimate their model under the assumption that the NAIRU is a random walk. Notice that in a sample of maximum 220 observations the two processes are basically indistinguishable. Furthermore, instead of assuming that policymakers fix the variance of the innovations to the NAIRU, they only have a prior on this variance and estimate it by maximizing the posterior. Assuming that they have such a prior is important for two reasons. First, it is realistic that, especially in small samples they tend to combine the prior and the likelihood information. Second, some prior is necessary in order to avoid misbehaviors of the likelihood like, for example, the fact that the likelihood might have a point mass at zero. This is known as the pile-up problem (see, for instance, Stock and Watson 1998).

As argued in section 2, $\hat{c}_\pi$, $\hat{c}_u$ and $\hat{u}_t^N$ are not separately identified from each other in the estimation of equations (7) and (8). However, this identification problem concerns only the level of the NAIRU and not its evolution over time, given that $\hat{c}_\pi$, $\hat{c}_u$ are time invariant. This means that the identification problem disappears if policymakers have a non-flat prior on the initial level of the NAIRU, which seems a reasonable assumption.

Summarizing, policymakers could potentially use all their information in a model consistent way, as long as they have a prior distribution on the initial level of the NAIRU and on the amount of time variation of the NAIRU.

Clearly, there exist priors that would have avoided most of the mismeasurement of the NAIRU, like very tight priors around a high level of the NAIRU in 1948. However, this seems rather unrealistic. It turns out instead that very sensible priors generate an evolution of policymakers' beliefs very similar to the one of figure 3, obtained with substantially more naive estimation techniques.

I assume that policymakers have a normal prior on $\hat{u}_0^N$ and an inverse-gamma (IG) prior on the variance of the innovations to the NAIRU, $\hat{\sigma}_{\tau}^{2}$. As an illustrative (and realistic) example, I assume that the prior on $\hat{u}_{0}^{N}$ has mean equal to 4.5 and variance equal to 1. Notice that 4.5 is a very plausible value as a prior mean for the NAIRU in 1948. The average unemployment rate between 1948 and 1958 was in fact 4.26. I also assume that the prior on $\hat{\sigma}_{\tau}^{2}$ has scale 0.04 and one degree of freedom. For the readers less familiar with the parametrization of the inverse-gamma distribution, an $IG(0.04, 1)$ distribution is relatively disperse (has such a fat right tail that it does not have a mean) and peaks at $\frac{0.04}{3}$ (which implies a standard deviation of the NAIRU in forty quarters of approximately 0.75, on top of the uncertainty on the initial level). Given such prior distribution on the NAIRU in 1948 and on its time variation, policymakers can estimate (7) and (8) by maximum likelihood using a Kalman filter and a Kalman smoother to get a posterior estimate of the states (the NAIRU). As before, they use their estimates as true values in the linear-quadratic policy problem, in order to choose the optimal level of the policy variable V.[28] Figure 14 plots the real-time estimates of the NAIRU, inflation persistence and the slope of the Phillips curve obtained in this way and compares them to the ones computed with the constant gain algorithm of section 2.3.1. The similarities between the results obtained with the two estimation methods are remarkable. As expected, the only difference is that the slope of the Phillips curve estimated with the maximum likelihood procedure remains equal to zero for a longer period of time, because the updating process is slower than with the constant gain algorithm.

Summarizing, the misperception of the NAIRU seems much more related to the unreliability of the end-of-sample estimates than to unsophisticated policymakers behavior. This point has been made before, by Orphanides and Van Norden (2001), amongst other authors. To conclude, the results of this subsection seem to suggest that relatively sophisticated policymakers would have equally misperceived the level of the NAIRU and the value of the other parameters of the economy. Consequently they would have not been able to avoid the outburst of inflation in the 60s and 70s.

7 Why Did Inflation Rise? Evidence from a Time Varying Phillips Curve

In the previous sections I have highlighted an important channel to understand and explain the low frequency behavior of the inflation and unemployment rates in the United States in the last forty years. I have argued that policymakers' mistakes in the estimation of the NAIRU explain why inflation started to increase in the early 60s. On the other hand, it was also made clear that the optimistic view about the level of the NAIRU is not enough to understand why rational policymakers let inflation rise for more than fifteen years. What is crucial to explain the dimension and the duration of the Great Inflation are policymakers' beliefs about the parameters of the Phillips curve. In particular, in the 60s policymakers did not fight inflation strongly enough because they underestimated the persistence of inflation in the Phillips curve. This induced the peculiar dynamics

[28] To speed up computations I do not estimate the parameters of the policy loss function (k and ϕ). Instead I fix them to the estimated value for the benchmark case.

described in section 3.2.

This being said, the important question becomes whether this kind of episodes can happen again in the future. As section 6.2 and other studies have argued, the mistakes in the real-time estimates of the NAIRU are more related to the unreliability of the end-of-sample estimates than to unsophisticated estimation algorithms. Therefore, there is no reason to think that the mistakes in the estimation of the NAIRU will be avoided in the future. However, the main insight of this paper is that economic stability is in danger when the misperception of the NAIRU is coupled with mistaken beliefs about the parameters of the Phillips curve. Therefore, it seems relevant to investigate whether big biases in the estimates of the Phillips curve can arise again.

Guessing what is likely to happen in the future depends crucially on understanding what did happen in the past, i.e. why policymakers in the 60s and early 70s made those mistakes that ultimately led to an inflation rate and an unemployment rate above 10%. At a deeper level, the important question becomes why policymakers in the 60s and 70s had such downward biased estimates of the persistence of inflation in the Phillips curve. Observe that the small sample bias does not seem to play an important role here. In fact, even with only twelve years of data, the probability of obtaining an estimate of $\alpha(1)$ smaller or equal than 0.5 is smaller than 2%, if the true Phillips curve is given by (5).[29] However, there are at least two alternative explanations of why $\hat{\alpha}(1)$ was so low in the 60s and early 70s. It is important to distinguish between them, because they might have rather different implications.

The first possible explanation is that, in the late 50s or early 60s, there was a break in the parameters of the Phillips curve. It is in principle possible that until the late 50s inflation persistence in the Phillips curve was lower than unity, implying an exploitable inflation-unemployment trade-off. This might have biased the estimates also in many of the subsequent time periods. If the Phillips curve (5) is interpreted as a reduced form relation, one possibility is that this break reflects a change in the policy regime. One option, for example, would be that the commitment of the dollar to the Gold Standard was taken more seriously during the 50s, while, instead, this was already not an issue for policymakers since the beginning of the 60s. As argued by Alogoskoufis and Smith (1991) and Alogoskoufis (1992), this would justify the low persistence of inflation in the immediate postwar period. According to this view, inflation rose because policymakers switched to a policy regime they were not used to. If this was the case, evaluating whether inflation is likely to rise again in the future would depend on how likely future policy regime changes are considered to be.[30] On the other hand, it is also possible that breaks in the Phillips curve reflect changes in some other structural parameters of the model, rather than changes in the policy regime. Under these circumstances, future similar breaks would be even less predictable.

[29] This probability can be calculated numerically, integrating out the nuisance parameters of the model.

[30] To be more precise, the main risk of high inflation would correspond to a switch from a policy regime implying low inflation persistence to another one associated with high persistence. For example, some authors (Benati 2003 or Debelle 2003) have argued that the persistence of inflation in the Phillips curve is lower under inflation targeting regimes. This would imply that abandoning inflation targeting could result in a rise of inflation persistence and, consequently, a dangerously weak policy reaction to inflation.

On the contrary, the second possible explanation is that there were no breaks in the inflation-unemployment dynamics. Nevertheless, policymakers ended up with biased estimates of the Phillips curve because of the "craziness" of the 50s (with the Korean war and the related price control periods) and/or because of the flatness of their priors on the structure of the economy (due to the lack of a solid economic ground). If this was the case, it is plausible to expect that both academics and policymakers have learned the lesson by now. While the misperception of the NAIRU is, to some extent, unavoidable, more developed economic theories and econometric techniques would prevent the other conditions that favored the Great Inflation. This would lead to a more optimistic conclusion.

In order to distinguish between these two working hypothesis, I estimate a modified version of the Phillips curve (5) on the sample 1948:I - 2002:IV. Here, I allow for time variation in the parameters and the volatility of the shocks. I do so by estimating the following Markov switching model, in which both parameters and volatility are allowed to switch among different regimes:

$$\pi_t = \bar{\pi}_t + \sigma_\zeta(S_t^\zeta)\zeta_t, \tag{23}$$

$$\bar{\pi}_t - \mu = b(S_t^P)\left[\alpha_1(\bar{\pi}_{t-1} - \mu) + \alpha_2(\bar{\pi}_{t-2} - \mu)\right] - \theta(L)(u_{t-1} - u_{t-1}^N) + \sigma_\xi(S_t^\xi)\xi_t, \tag{24}$$

where ζ_t and ξ_t are *i.i.d.* $N(0,1)$ and u_t^N follows the exogenous stochastic process specified in (3). μ is a parameter representing the level of average inflation when unemployment is at its NAIRU and the sum of coefficients on past inflation is not equal to one. $b(S_t^P)$ is a switching parameter, meant to capture the scale of the sum of coefficients on past inflation. Notice that, being mainly interested in breaks in persistence and in order to reduce the number of free parameters, the slope of the Phillips curve and the constant μ are not allowed to change over time. This is consistent with the findings of Staiger, Stock, and Watson (2001). $\sigma_\zeta(S_t^\zeta)$ and $\sigma_\xi(S_t^\xi)$ represent the standard deviations respectively of the shocks ζ_t and ξ_t. These standard deviations are allowed to switch as well among different states.

Notice that (23) and (24) represent a model in which inflation consists of two components: a signal component $\bar{\pi}_t$, corresponding to the "theoretical" inflation rate and an *i.i.d.* component ζ_t, interpretable as measurement error or just *noise*. If, relative to the signal, the noise component is large, the estimate of equation (5) would clearly produce downward biased estimates of α. One can think of the 50s as a period in which the variance of the noise was high enough to importantly affect the real-time estimates of (5). In other words, this modification of the standard Phillips curve is important to investigate the plausibility of the second proposed explanation of the bias in the estimates of inflation persistence.

I will consider several possible models, which differ in the number of regimes allowed for parameters and variances. The complete set of estimated models is listed in table 7, which also reports three measures of the models' fit. The first three are models in which the variance of the noise component is constrained to be equal to zero. In this case, the estimated equations closely resemble the standard Phillips curve of equation (5). The difference is that here the variance of the innovations to (24) switches among three possible states. For the parameter representing inflation persistence I experiment with one, two or three possible regimes. The second set of three models is characterized by two states for both the variances (of the noise and the innovations to the signal in (24)) and again, one, two or three states for the persistence parameter. In the last set of three models, instead I set the variance of the noise component to be equal to zero in one of the two regimes.

To increase flexibility I assume that parameters and volatilities switch independently with transition probabilities given by

$$\begin{aligned}
\Pr(S_t^P &= i|S_{t-1}^P = j) = q_{ij}^P \\
\Pr(S_t^\zeta &= i|S_{t-1}^\zeta = j) = q_{ij}^\zeta \\
\Pr(S_t^\xi &= i|S_{t-1}^\xi = j) = q_{ij}^\xi .
\end{aligned}$$

When I allow for three distinct regimes for the disturbance volatilities, I impose the following restrictions on the matrix of transition probabilities:

$$Q^V = \begin{bmatrix} q_1^V & 1-q_1 & 0 \\ \frac{1-q_2^V}{2} & q_2^V & \frac{1-q_2^V}{2} \\ 0 & 1-q_3^V & q_3^V \end{bmatrix}.$$

The structure of Q^V reflects the assumption of adjacent switching (Sims 1999) and, again, is motivated by the necessity to reduce the number of free parameters. In the model with three regimes for the parameters, the matrix of transition probabilities for the parameters has a structure similar to Q^V. In the models with two regimes the matrix of transition probabilities is left unrestricted.[31]

To estimate the models I maximize the joint posterior of the models' free parameters (switching coefficients, constant coefficients, switching variances and transition probabilities). The likelihood of these models can be written following Hamilton (1994), Sims (1999 and 2001) or Kim and Nelson (1999b). The posterior is obtained multiplying the likelihood by the priors. As pointed out in section 4.1.2, considering the priors is crucial for model comparison. I use normal priors for parameters, inverse-gamma for variances and beta for the transition probabilities. They are standard and reasonably flat on the interesting regions of the parameter's space. The complete list of priors is reported in table 6.

According to both the BIC and the marginal likelihood (table 7), the best fitting model is the model with just one regime for the inflation persistence parameter, two regimes for the volatility of the innovation to the signal (ξ_t) and two regimes for the noise (ζ_t). The estimates (posterior mode and standard deviation) of the best fitting model are reported in table 8, which also reports the likelihood mode for comparison. Notice that, in this model, the sum of coefficients on past inflation is equal to 0.95, which is very close to the theoretical value of no long run inflation-unemployment trade-off. Furthermore, while the first regime for the volatility of the noise component is a low

[31] Notice that a model with three states for the parameters and three states for the innovation volatility has nine total states. As an example, in this case the complete transition probabilities matrix would be given by $Q^P \otimes Q^V$.

volatility regime, the second is a high volatility one. In fact, the standard deviation of the noise is equal to 4.097 in the second regime. Figure 15 plots the posterior probability of the high volatility regime for the noise component. Figure 15 also reports the inflation rate, rescaled for comparison. It is easy to see that this high volatility regime occurs only in the first part of the sample, from 1948 to 1953. This is exactly the period of time when the behavior of inflation was affected by the end of World War II, the Korean war and the decisions about price and wage controls related to the two wars. This offers a plausible explanation for the downward bias of the real-time estimates of the inflation persistence in the Phillips curve in the 60s and 70s.

Looking also at the other models and not only at the best fitting one, two conclusions emerge from the analysis of table 7. First, all models with three regimes for the inflation persistence parameter perform quite poorly. Second, all models with one or two regimes for the inflation persistence parameter have measures of fit, which are relatively close to the ones of the best fitting model. However, analyzing the estimates (not reported) of the models with two regimes for the persistence, it is remarkable that the value of the inflation persistence is never below 0.73, which is not low enough to explain the real-time estimates of the 60s.

Concluding, while the explanation based on a break in the inflation persistence parameter cannot be completely ruled out, the data seem to support the hypothesis that policymakers in the 60s had downward biased estimates of the persistence of inflation in the Phillips curve because they did not recognize the large amount of noise around the main path of the inflation rate, which characterized the 50s.

8 Concluding Remarks

This paper presents a simple model of inflation-unemployment dynamics when policymakers have imperfect information. The source of imperfect information is the fact that policymakers do not know the exact model of the economy. Therefore, they update their beliefs about the model's unknowns in every period and implement optimal policy, conditional on their current beliefs.

The model's self-confirming equilibrium is characterized by low inflation. Nevertheless, the model can generate prolonged and asymmetric episodes of high inflation, which closely resemble the run-up of US inflation in the 60s and 70s and the sharp disinflation in the early 80s. In particular, these episodes occur when policymakers simultaneously underestimate *both* the NAIRU *and* the persistence of inflation in the Phillips curve. Starting from this situation of overoptimistic beliefs, the model endogenously generates periods in which policy is overpessimistic. During these periods policy mistakenly perceives the inflation-unemployment trade-off as too costly. I show that it is not optimal to create a recession to stop inflation, either if beliefs are overoptimistic or if they are overpessimistic. This is why the policy action against inflation is delayed until the moment in which the perceived trade-off improves. When this happens, inflation is already very high. Therefore, the policy action against inflation is strong and decisive.

Unlike most of the existing literature, the model matches many recognized stylized facts of the

American Great Inflation of the 60s, 70s and 80s. I also formally estimate the model by likelihood methods and find that it fits the behavior of US inflation and unemployment remarkably well.

Given the empirical support, the model can be used to evaluate the possibility that episodes similar to the Great Inflation could happen again in the future. In this respect, the conclusion of the paper is more optimistic than alternative theories. In this model, high inflation episodes do not occur as the consequence of the mismeasurement of the NAIRU only (see, for instance, Orphanides 2000) or as a result of policymakers' lack of commitment to low inflation (see, for instance, Sargent 1999). High inflation is the consequence of the unlikely combination of two factors. As mentioned above, one of these two factors is the mismeasurement of the NAIRU, which, to some extent, is unavoidable. However, the serious underestimation of the persistence of inflation in the Phillips curve seems to be much more uncommon. As argued in section 7, it appears more related to special circumstances, such as structural breaks in the true model of the economy or, as in the case of the Great Inflation, periods of extremely high volatility of some temporary disturbances.

A Self-confirming equilibrium

This appendix formally defines a self-confirming equilibrium and derives the self-confirming equilibrium of the model of section 2, in the case of constant gain learning. Let y_t^π, y_t^u, x_t^π, x_t^u, $\hat{\beta}^\pi$ and $\hat{\beta}^u$ be the same objects defined in section 2.3.1. I follow Sargent (1999) in defining a self-confirming equilibrium in this framework:

Definition 1 *A self-confirming equilibrium is a set of policymakers' beliefs about the models' parameters $\hat{\beta} \equiv [\hat{\beta}^\pi, \hat{\beta}^u, \hat{u}^N]$, a fixed optimal policy rule $g(\hat{\beta})$ and an associated stationary stochastic process for the vector $\left[\pi_t, u_t, V_t, u_t^N\right]$ such that: (a) $\hat{u}^N$, $\hat{\beta}^\pi$ and $\hat{\beta}^u$ satisfy*

$$E\left[u_t - \hat{u}^N\right] = 0 \tag{25}$$

$$E\left[x_t^i\left(y_t^i - x_t^{i\prime}\hat{\beta}^i\right)\right] = 0, \qquad i = \{\pi, u\} \tag{26}$$

where the expectations are taken with respect to the probability distribution generated by (5), (2), (3) and (9); (b) the vector $\left[\pi_t, u_t, V_t, u_t^N\right]$ is generated by the stationary stochastic process implied by (5), (2), (3) and (9).

It is straightforward to verify that the set of beliefs $\hat{u}^N = u^*$, $\hat{\beta}^\pi = [0; \alpha_1; \alpha_2; \theta_1; \theta_2]$ and $\hat{\beta}^u = [0; \rho_1; \rho_2]$ satisfy (25) and (26) and therefore represents a self-confirming equilibrium, in the case of $\sigma_\tau^2 = 0$. When $\sigma_\tau^2 > 0$, finding the self-confirming equilibrium is more involved and requires a numerical solution of the system of equations given by (25) and (26). The procedure works as follows: any given fixed value of $\hat{u}^N$, $\hat{\beta}^\pi$ and $\hat{\beta}^u$ implies a linear stochastic process for $\left[\pi_t, u_t, V_t, u_t^N\right]$ via equations (5), (2), (3) and (9). The linear process can be rewritten as a first order system of the form $z_t = C + Az_{t-1} + B\nu_t$. Thus $E(z_t) = (I - A)^{-1}C$ and $Var(z_t)$ can be found by solving the Lyapunov equation $Var(z_t) = AVar(z_t)A' + BVar(\nu_t)B'$. The elements of $E(z_t)$ and $Var(z_t)$ can be used to compute $E\left[u_t - \hat{u}^N\right]$ and $E\left[x_t^i(y_t^i - x_t^{i\prime}\hat{\beta}^i)\right]$, for $i = \{\pi, u\}$, which, in general, will not be equal to zero. A simple equation solver can be used to solve for the set of beliefs $\hat{u}^N$, $\hat{\beta}^\pi$ and $\hat{\beta}^u$, which satisfy (25) and (26). Of course the solution will depend on the model's true parameters. As an illustrative example I consider the case in which the model's true parameters are the point estimates of the baseline model, presented in the first column of table 1. The self-confirming equilibrium in this case corresponds to $\hat{u}^N = 6$, $\hat{\beta}^\pi = [0.0394; 0.7203; 0.2623; -0.8409; 0.7637]$ and $\hat{\beta}^u = [0; 1.5703; -0.6269]$. Furthermore, the eigenvalues of the Jacobian of the expressions contained in (25) and (26), evaluated at the self-confirming equilibrium, have negative real parts. This guarantees the stability of the equilibrium.

As mentioned in section 2.4, as $Var\left(u_t^N\right)$ increases the mistakes associated with the estimation of the current level of the NAIRU will bias toward zero the estimate the slope of the Phillips curve and of the persistence of unemployment deviations from the NAIRU in the aggregate demand equation. This leads to self-confirming equilibria whose distance from the true parameters' values is increasing in $Var\left(u_t^N\right)$. Finally, when $Var\left(u_t^N\right)$ is large with respect to σ_ε^2 and σ_η^2, the model does not admit a self confirming equilibrium anymore. Figure 16 plots the Euclidean distance between the true parameters ($[0; \alpha_1; \alpha_2; \theta_1; \theta_2; 0; \rho_1; \rho_2; u^*]$) and the set of beliefs about these parameters corresponding to the self-confirming equilibria. These self-confirming equilibria are computed for different values of $Var\left(u_t^N\right) \geq 0$. This graph confirms the intuition that the distance between equilibrium beliefs and true parameters increases with $Var\left(u_t^N\right)$. The line is truncated at the value 4.63, because for $Var\left(u_t^N\right)$ larger than this value a self confirming equilibrium cannot be found. All the found self-confirming equilibria are stable.

B State space form for model's estimation

This appendix gives the details of the state space form representation of the model for the estimation with the Kalman filter.

The canonical state space form is given by:

$$y_t = AZ_t + BX_t + Re_t, \tag{27}$$

$$X_t = C + GX_{t-1} + Qs_t, \tag{28}$$

$$\begin{pmatrix} e_t \\ s_t \end{pmatrix} \sim i.i.d.\ N(0, I). \tag{29}$$

In our case, $y_t = [\pi_t; u_t]'$; $Z_t = [\pi_{t-1}; \pi_{t-2}; u_{t-1}; u_{t-2}; V_{t-1}]'$; $X_t = \left[u_t^N; u_{t-1}^N; u_{t-2}^N\right]'$;

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & -\theta_1 & -\theta_2 & 0 \\ 0 & 0 & \rho_1 & \rho_2 & 1 \end{bmatrix};\ B = \begin{bmatrix} 0 & \theta_1 & \theta_2 \\ 1 & -\rho_1 & \rho_2 \end{bmatrix};\ C = \begin{bmatrix} (1-\gamma)u^* \\ 0 \\ 0 \end{bmatrix};\ G = \begin{bmatrix} \gamma & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix};$$

$$R = \begin{bmatrix} \sigma_\varepsilon & 0 \\ 0 & \sigma_\eta \end{bmatrix};\ Q = \begin{bmatrix} \sigma_\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The standard Kalman filter recursion formulas can be found in Hamilton (1994). To start the recursion it is necessary to specify $E\left(X_0|\Omega_0\right)$ and $Var\left(X_0|\Omega_0\right)$, where Ω_0 represents the information set available at time 0. Following a common practice, I set $E\left(X_0|\Omega_0\right)$ and $Var\left(X_0|\Omega_0\right)$ to the unconditional values implied by the transition equation. In particular, this results in $E\left(X_0|\Omega_0\right) = [6; 6; 6]'$, which corresponds approximately to the estimate of the NAIRU of Staiger, Stock, and Watson (2001) in 1960 (which is the initial date of our sample).

C Marginal likelihood

This appendix illustrates how to compute the marginal likelihood for the VARs, in order to perform the model comparison. As mentioned in section 4.1.2, the marginal likelihood for the learning models is computed using the Laplace approximation method, described in details in Schervish (1994) or Sims (2002). Instead of relying on approximations, for systems of normal regression equations (with natural conjugate priors) it is possible to derive the exact formula of the marginal likelihood.

Notice that VAR models belong to this category, if we condition on the initial observation, which is standard in the literature.

Consider the generic system of linear equations

$$y_t = BX_t + \varepsilon_t,$$

where y_t is an $n \times 1$ vector of observed endogenous variables; X_t is an $k \times 1$ vector of observed predetermined variables; B is an $n \times k$ matrix of unknown coefficients; ε_t is an $n \times 1$ vector of unobservable shocks, distributed *i.i.d.* $N(0, \Omega)$. Let $\theta \equiv [B; \Omega]$, i.e. the set of unknown parameters of the model, $y \equiv [y_1, ..., y_T]'$ and $X = [X_1, ..., X_T]'$. The marginal likelihood is defined as

$$m(y) = \int f(y|\theta)\, \pi(\theta)\, d\theta,$$

where the integral is taken with respect to the entire support of θ, $f(\cdot)$ is the likelihood function and $\pi(\cdot)$ is the prior. Under natural conjugate prior, the kernels of the posterior density and the likelihood function have the same functional form. Therefore, the key step to obtain the exact marginal likelihood is to compute the following integral of the normal-inverse-Wishart density kernel (where, only for simplicity, I am neglecting the effect of the prior):

$$m(y) = \int \int (2\pi)^{-\frac{nT}{2}} |\Omega|^{-\frac{T}{2}} \underbrace{\exp\left(-\frac{1}{2}\left[\sum (y_t - BX_t)\right]' \Omega^{-1} \left[\sum (y_t - BX_t)\right]\right)}_{A} dBd\Omega \tag{30}$$

Observe that A can be rewritten as

$$\exp\left(-\frac{1}{2}\left[\sum \left(y_t - \hat{B}X_t\right)\right]' \Omega^{-1} \left[\sum \left(y_t - \hat{B}X_t\right)\right] - \frac{1}{2}\left(\hat{B} - B\right)' \left[(X'X) \otimes \Omega^{-1}\right] \left(\hat{B} - B\right)\right), \tag{31}$$

where $\widehat{B}$ is the value of B that maximizes the posterior.[32] Observe that (31) is a quadratic kernel in B. Therefore, substituting (31) into (30) and taking the integral with respect to B, we obtain

$$m(y) = \int (2\pi)^{-\frac{nT-nk}{2}} |\Omega|^{-\frac{T-k}{2}} \left|X'X\right|^{-\frac{n}{2}} \underbrace{\exp\left(-\frac{1}{2}\left[\sum \left(y_t - \hat{B}X_t\right)\right]' \Omega^{-1} \left[\sum \left(y_t - \hat{B}X_t\right)\right]\right)}_{D} d\Omega,$$

Let $S \equiv \left[\sum \left(y_t - \hat{B}X_t\right)\right] \left[\sum \left(y_t - \hat{B}X_t\right)\right]'$. Then D can be rewritten as $\exp\left(-\frac{1}{2}tr\left(\Omega^{-1}S\right)\right)$. Observe that

$$|\Omega|^{-\frac{T-k}{2}} \exp\left(-\frac{1}{2}tr\left(\Omega^{-1}S\right)\right)$$

is the kernel of an inverse-Wishart distribution in Ω, with degrees of freedom $v = T - k - n - 1$ and scale matrix S. Thus Ω can be integrated out easily, leading to the final expression

$$m(y) = (2\pi)^{-\frac{nT-nk}{2}} \left|X'X\right|^{-\frac{n}{2}} C,$$

where

$$C = 2^{\frac{1}{2}vn} \pi^{\frac{1}{4}n(n-1)} \prod_{i=1}^{n} \Gamma\left(\frac{v+1-i}{2}\right) |S|^{-\frac{1}{2}v}.$$

[32] In a reduced form VAR with flat or conjugate priors implemented through a training sample or dummy observations, $\hat{B}$ are the OLS estimates of the coefficients.

D The solution method for the forward looking model

This appendix illustrates in more detail the method used to solve the forward looking model with fully rational agents. As mentioned in section 5.2, the model is hard to solve because it is a nonlinear system of rational expectation equations. The source of nonlinearity is the learning behavior of policymakers. I will rely on numerical methods. The adopted solution method is based on Fackler and Miranda (2001). A similar method is in Fernandez-Villaverde and Rubio (2002).

Consider the system of rational expectation equations given by (19) and (20). To simplify the analysis and only for the purposes of this section I will assume that u_t^N is a deterministic function, known by the private sector, but, as usual, unknown by policymakers. I will set u_t^N to be equal to the smoothed estimate of u_t^N obtained in the estimation of the forward looking model with partially rational agents (section 5.1). Thus, let $\tilde{u}_t \equiv u_t - u_t^N$. Equations (19) and (20), the only ones involving expectations, can be rearranged and rewritten in the following compact form:

$$y_t = AE_{t-1}y_{t+1} + BX_{t-1} + v_t, \tag{32}$$

where $y_t \equiv [\pi_t, \tilde{u}_t]'$ is the vector of observed endogenous variables; $X_{t-1} \equiv [\pi_{t-1}, \tilde{u}_{t-1}, V_{t-1}]$ is the vector of observed predetermined variables; $v_t \equiv [\varepsilon_t, \eta_t]'$ is the vector of unobservable shocks; A and B are matrices of coefficients, omitted for brevity. (32) is linear, but the complete system, given by (32), (9), (10), (11), (12) and (13) is nonlinear. The solution of the model is the unknown response function $E_{t-1}y_{t+1} = \Psi(\Omega_{t-1})$, where Ω_t represents the information available at time t and $\Psi(\cdot)$ satisfies

$$\Psi(\Omega_{t-1}) = AE_{t-1}\Psi(\Omega_t) + BE_{t-1}X_t.$$

When the model is nonlinear in general there is not a closed form expression for $\Psi(\cdot)$ and it must be approximated numerically by projection methods. The basic idea of Fackler and Miranda (2001) is approximating $\Psi(\Omega_t)$ with a linear combination of basis functions of the state variables. This is given by Φs_t, where s_t is an $m \times 1$ vector of basis functions and Φ is a $2 \times m$ matrix of coefficients. In the numerical procedure, also the expectation operator must be approximated using quadrature methods. Therefore, the expectation of a generic function $f(\cdot)$ of the model's source of randomness, v_t, is approximated by a discrete version of the integral, given by

$$Ef(v_t) \approx \sum_{j=1}^{k} \omega_j f\left(v_t^j\right).$$

For a given value of the innovations v_t^j,

$$y_t^j = A\Phi s_{t-1} + BX_{t-1} + v_t^j$$

and

$$y_{t+1}^j = A\Phi s_t^j + BX_t + v_{t+1},$$

where the superscript j for s indicates that the value of s at time t depends on the realization of the shocks at time t. Now we can compute $E_{t-1}y_{t+1} = E_{t-1}E_t y_{t+1} \approx \sum_{j=1}^{k} \omega_j \left(A\Phi s_t^j + BX_t \right) \equiv z_{t-1}$. Let $S \equiv [s_{1,t-1}, ..., s_{n,t-1}]$ be a collection of $n \geq m$ values of s_{t-1} and $Z \equiv [z_{1,t-1}, ..., z_{n,t-1}]$ the collection of the corresponding n values of z_{t-1}. The solution consists in the Φ^* which solves $\Phi^* S = Z$, or $\Phi^* SS' = ZS'$ in the case in which $n > m$. It can be done using standard equation solvers.

In this application, to approximate the integrals and expectation operators, I use a Gauss-Hermite quadrature with 3 nodes (see Judd 1998). As mentioned above, the dimension of the state vector is high. Thus, the use of tensor product bases or complete polynomial bases is unfeasible for any polynomial degree bigger or equal to 2. For this reason I chose the following ad hoc collection of 21 basis functions of the states, which turned out to work well:

$$
\begin{aligned}
s_t \quad = \quad & \Big[1; \pi_t; \tilde{u}_t; \pi_{t-1}; \tilde{u}_{t-1}; V_t; \hat{c}_{\pi,t}; \hat{\alpha}_{1,t}; \hat{\alpha}_{2,t}; \hat{\theta}_{1,t}; \hat{\theta}_{2,t}; \hat{c}_{u,t}; \hat{\rho}_{1,t}; \hat{\rho}_{2,t}; \hat{u}_t^N; \\
& \pi_t^2; \tilde{u}_t^2; \hat{\alpha}_{1,t}\pi_t; \hat{\alpha}_{2,t}\pi_{t-1}; \hat{\theta}_{1,t}\left(u_t - \hat{u}_t^N\right); \hat{\theta}_{1,t}\left(u_{t-1} - \hat{u}_{t-1}^N\right) \Big].
\end{aligned}
$$

Notice that the choice of s_t includes all linear terms in the state variables of the problem and some potentially relevant second order terms. The dimension of s_t is so large that the choice of S based on standard grid methods is unfeasible, even specifying only two values for any state variable. To solve this problem I chose a collection of $n = 86$ s_t's, corresponding to the actual values of s_t observed in the data, every 2 quarters, from 1959:IV to 2002:IV. The results are only marginally affected by a different choice of values for S, like for example the observed data, every 2 quarters, 1960:I to 2002:III.

E The MCMC algorithm for the stochastic volatility model

This appendix illustrates the details of the MCMC algorithm used in section 6.1 for the estimation of the model with stochastic volatility. The parameters of interest are the coefficients $\Psi_1 \equiv [\alpha_1; \theta_1; \theta_2; \rho_1; \rho_2]$, $\Psi_2 \equiv [k; \phi]$, $\Psi_3 \equiv \left[\sigma_{\nu_\varepsilon}^2; \sigma_{\nu_\eta}^2\right]$ and the unobservable states $u^N \equiv \left\{u_t^N\right\}_{t=1}^T$, $\sigma_\varepsilon \equiv \{\sigma_{\varepsilon,t}\}_{t=1}^T$ and $\sigma_\eta \equiv \{\sigma_{\eta,t}\}_{t=1}^T$. The estimation consists of the simulation of the posterior of the parameters of interest, conditional on the observed data. MCMC allows to simulate lower dimensional conditional posteriors instead of the high dimensional unconditional one.

Notice that the model can be rewritten like in (27), (28) and (29), with the difference that now the elements of R are time varying and follow the processes (21) and (22). The algorithm works in 5 steps.

E.1 Step 1: drawing u^N

Conditional on σ_ε, σ_η, Ψ_1 and Ψ_2, the observation equation (27) is linear and has Gaussian innovations with known variance. Therefore, the vector u^N can be drawn using standard simulation smoothers, like, for instance, Carter and Kohn (1994) or Durbin and Koopman (2002). Details of this procedure can be also found in Kim and Nelson (1999b).

E.2 Step 2: drawing σ_ε and σ_η

Consider now the system of equations

$$y_t - AZ_t - BX_t = y_t^* = R_t e_t \tag{33}$$

where, taking u^N, Ψ_1 and Ψ_2 as given, y_t^* is observable. This is a system of nonlinear measurement equations, but can be easily converted in a linear one, by squaring and taking logarithms of every element of (33), which leads to the following approximating state space form:

$$y_t^{**} = 2h_t + e_t^{**} \tag{34}$$

$$h_t = h_{t-1} + \omega_t. \tag{35}$$

$y_{it}^{**} = \log[(y_{it}^*)^2 + \bar{c}]$; $\bar{c}$ is an offset constant (set to 0.001); $e_{it}^{**} = \log(e_{it}^2)$; $h_t = \log(\operatorname{diag}(R_t))$. Observe that the e^{**}'s and the ω's are not correlated. The system in this form has a linear, but non-Gaussian state space form, because the innovations in the measurement equations are distributed as a $\log\chi^2(1)$. In order to further transform the system in a Gaussian one, a mixture of normals approximation of the log χ^2 distribution is used, as described in Kim, Shephard, and Chib (1998). Observe that the variance covariance matrix of the e's is the identity matrix. This implies that the variance covariance matrix of the e^{**}'s is also diagonal, allowing to use the same (independent) mixture of normals approximation for any element of e^{**}. Kim, Shephard and Chib (1998) select a mixture of 7 normal densities with component probabilities q_i, means m_i, and variances v_i^2, $i = 1, ..., 7$. The constants $\{q_i, m_i, v_i^2\}$ are chosen to match a number of moments of the $\log\chi^2(1)$ distribution. The constants $\{q_i, m_i, v_i^2\}$ can be found in Kim, Shephard, and Chib (1998).[33]

For the innovation to the variable y_{jt}, define as $s_j^T = [s_{j1}, ..., s_{jT}]'$ the vector of indicator variables selecting at every point in time which member of the mixture of normal approximation has to be used. Conditional on u^N, Ψ and s^T (which denotes the collection of s_j^T), the system has an approximate linear and Gaussian state space form. Again, exactly like in the previous step of the sampler, this procedure allows to draw every h_t using a simulation smoother.

Conditional on the data and the new series of h_t's, it is possible to sample the new s_j^T vectors, to be used in the next iteration. This is easily done (separately for every j) by sampling from the discrete densities defined by

$$\Pr(s_{jt} = i \mid y_{jt}^{**}, h_{jt}) \propto q_i f_N(y_{jt}^{**} \mid 2h_{jt} + m_i, v_i^2), \quad i = 1, ..., 7.$$

Further details can be found in Kim, Shephard, and Chib (1998) or Primiceri (2003).

[33] In this paper the reweighting procedure used in Kim, Shephard and Chib (1998) to correct the minor approximation error is not adopted.

E.3 Step 3: drawing Ψ_1

Conditional on Ψ_2, σ_ε, σ_η and u^N, the objects Z_t, X_t and R_t are observable. Therefore, the elements of Ψ_1 (which correspond to the elements of A and B) can be easily drawn from the posterior of the coefficients of a regression with known variance. This posterior is normally distributed with mean equal to the OLS coefficients and variance equal to the variance of the OLS coefficients.

E.4 Step 4: drawing Ψ_2

Ψ_2 enters the model non-linearly. Therefore, in order to draw from the conditional posterior of Ψ_2, I use a Metropolis step, nested in the Gibbs sampler. The procedure works as follows: I draw a candidate value Ψ_2^* from a proposal distribution $\varphi\left(\Psi_2^*|\Psi_2^{i-1}\right)$, where Ψ_2^{i-1} is the previous draw of the chain. At this point I compute the value of the posterior associated to the draw, $p\left(\Psi_2^*|\Psi_1, \sigma_\varepsilon, \sigma_\eta, u^N, \{y_t\}_{t=1}^T\right)$, which, under flat prior, is proportional to the value of the likelihood. The new draw is accepted with probability

$$a = \min\left\{\frac{p\left(\Psi_2^*\right)/\varphi\left(\Psi_2^*\right)}{p\left(\Psi_2^{i-1}\right)/\varphi\left(\Psi_2^{i-1}\right)}, 1\right\}.$$

If the proposal value is rejected, the next element of the chain is set to be Ψ_2^{i-1}. In order to satisfy the constraints $\phi \geq 0$ and m, $0 \leq k \leq 1$, I chose the proposal distribution to be normal in $f(\Psi_2^*)$, where $f\left(a, b\right) = \left[\log(a), \log\left(\frac{b}{1-b}\right)\right]$. The mean is chosen to be $f(\Psi_2^{i-1})$, while I fix the variance to a diagonal matrix with elements 0.001 and 0.005 on the main diagonal.

E.5 Step 5: drawing Ψ_3

Conditional on σ_ε, σ_η, each element of Ψ_3 has an inverse-Gamma posterior distribution, independent of the other element. Conditional on σ_ε, σ_η, it is easy to draw from these inverse-Gamma posteriors because the innovations are observable.[34]

[34] See Gelman, Carlin, Stern, and Rubin (1995) for a description of the sampling procedure from an inverse-Gamma or inverse-Wishart distributions.

Coefficients	CG	OLS	DLS	SV
α_1	0.707 (0.074)	0.711 (0.074)	0.707 (0.073)	0.663 (0.076)
θ_1	-1.053 (0.292)	-1.021 (0.290)	-1.057 (0.291)	$-.718$ (0.244)
θ_2	0.928 (0.289)	0.903 (0.287)	0.943 (0.289)	0.63 (0.241)
ρ_1	1.661 (0.057)	1.756 (0.065)	1.640 (0.056)	1.376 (0.083)
ρ_2	-0.737 (0.057)	-0.779 (0.060)	-0.719 (0.056)	-0.449 (0.082)
σ^2_ε	1.033 (0.113)	1.041 (0.113)	1.033 (0.109)	–
σ^2_η	0.036 (0.006)	0.036 (0.006)	0.035 (0.006)	–
ϕ	2131 (1570)	475.5 (273.9)	1902 (1119)	2763 (2375)
k	0.872 (0.026)	0.960 (0.014)	0.809 (0.032)	0.869 (0.026)

Table 1: Maximum likelihood estimates of the model's parameters. CG: model with constant gain learning; DLS: model with discounted least squares learning; OLS: model with ordinary least squares learning; SV: model with stochastic volatility. Standard errors in parenthesis.

Models	log-Likelihood	BIC	log-Marginal Likelihood
CG	−250.16	−273.30	−273.66
OLS	−252.02	−275.16	−268.22
DLS	−248.99	−272.13	−263.68
VAR(2)	−242.74	−276.17	−278.70
VAR(3)	−234.52	−278.22	−279.79
VAR(4)	−229.90	−283.89	−279.98

Table 2: Measures of fit of different models. CG: model with constant gain learning; DLS: model with discounted least squares learning; OLS: model with ordinary least squares learning; VAR: vector autoregressions with 2, 3 of 4 lags.

Parameters	CG	FL_1	FL_2
$\max(\pi)$	11.95 [10.52; 13.44]	14.71 [12.87; 16.45]	13.90 [12.60; 15.27]
$\max(u)$	10.60 [9.31; 12.25]	11.16 [9.04; 13.42]	10.16 [8.89; 11.86]
t^*_π	1975:III [1971; 1981]	1975:II [1970:I; 1985:I]	1968 [1966; 1975:II]
$t^*_u - t^*_\pi$	2:III [2; 6:I]	4:II [2:II; 8:I]	6:I [4:I; 9:III]

Table 3: Summary statistics of the simulations (medians, lower and upper quartiles). CG: baseline model, with constant gain learning; FL_1: forward looking model with partially rational agents; FL_2: forward looking model with fully rational agents (t^*_π and t^*_u stand for the time periods in which inflation and unemployment peak).

Coefficients	SP	SP (no learning)
α_1	0.709 (0.073)	0.718 (0.074)
θ_1	-1.035 (0.295)	-1.102 (0.285)
θ_2	0.910 (0.291)	1.013 (0.278)
ρ_1	1.635 (0.062)	1.568 (0.066)
ρ_2	-0.724 (0.055)	-0.699 (0.062)
σ_ε^2	1.037 (0.111)	1.039 (0.114)
σ_η^2	0.034 (0.006)	0.030 (0.006)
ϕ	950.6 (583.4)	2764 (16871)
k_1	1.000 (0.000)	0.001 (0.001)
k_2	1.000 (0.007)	0.000 (0.054)
π_1^*	4.182 (0.739)	3.598 (2.020)
π_2^*	3.535 (0.345)	2.297 (1.138)
λ_1	0.847 (0.566)	1.706 (2.205)
λ_2	0.420 (0.270)	0.324 (1.109)
log-Likelihood	-245.99	-241.94
BIC	-281.98	-277.94

Table 4: Maximum likelihood estimates of the model's parameters. SP: model with constant gain learning and shift in the parameters representing the policy preferences; SP (no learning): model with shift in the parameters representing the policy preferences and without learning on the persistence of inflation and the slope of the Phillips curve. Standard errors in parenthesis. See section 4.3 for details.

θ	ρ_b	ρ_f	σ^2_ε	σ^2_η	ϕ	k	log-lh	BIC
−0.00095	0.936	0.0000	1.18	0.0833	1972	0.880	−307.53	−325.52
(0.00081)	(0.0237)	(0.0025)	(0.1739)	(0.0152)	(1396)	(0.034)		

Table 5: Maximum likelihood estimates, log-likelihood and BIC for the forward looking model with partially rational agents. Standard errors in parenthesis.

Parameters	Distribution	Mean	Standard Deviation
μ	N	2	0.75
α_1	normalized	0.6	–
α_2	N	0.4	0.75
θ_1	N	−0.7	0.75
$\theta(1)$	N	−0.1	0.75
$b\left(S_t^P = 1\right)$	N	1	0.75
$b\left(S_t^P = 2\right)$	N	0.75	0.75
$b\left(S_t^P = 3\right)$	N	0.5	0.75
$\sigma^2_\xi\left(S_t^\xi = 1\right)$	IG	1.33	1.89
$\sigma^2_\xi\left(S_t^\xi = 2\right)$	IG	3.33	4.71
$\sigma^2_\xi\left(S_t^\xi = 3\right)$	IG	33.33	47.14
$\sigma^2_\zeta\left(S_t^\zeta = 1\right)$	IG	1.33	1.89
$\sigma^2_\zeta\left(S_t^\zeta = 2\right)$	IG	33.33	47.14
$q^P_{ii}, \quad i = \{1,2,3\}$	$Beta$	0.98	0.98
$q^\xi_{ii}, \quad i = \{1,2,3\}$	$Beta$	0.98	0.98
$q^\zeta_{ii}, \quad i = \{1,2\}$	$Beta$	0.98	0.98

Table 6: Prior distributions used for the estimation and model comparison of the Markov switching models of section 7 (N and IG stand respectively for the Normal and Inverse-Gamma distributions).

Models	log-Lh	BIC	log-Marginal Lh
$n^P = 1, n^\xi = 3, n^\zeta = 1, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−330.38	−360.00	−348.91
$n^P = 2, n^\xi = 3, n^\zeta = 1, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−325.20	−362.89	−348.51
$n^P = 3, n^\xi = 3, n^\zeta = 1, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−325.42	−368.49	−351.16
$n^P = 1, n^\xi = 2, n^\zeta = 2$	−322.98	−357.98	−347.17
$n^P = 2, n^\xi = 2, n^\zeta = 2$	−316.15	−359.22	−347.82
$n^P = 3, n^\xi = 2, n^\zeta = 2$	−316.57	−365.03	−349.08
$n^P = 1, n^\xi = 2, n^\zeta = 2, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−329.31	−361.62	−349.60
$n^P = 2, n^\xi = 2, n^\zeta = 2, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−324.59	−364.98	−349.73
$n^P = 3, n^\xi = 2, n^\zeta = 2, \sigma^2_\zeta \left(S^\zeta_t = 1 \right) = 0$	−324.82	−370.58	−352.21

Table 7: Fit of the Markov switching models (n^P, n^ξ and n^ζ stand respectively for the number of regimes of the parameters, variance of the autocorrelated disturbances and variance of the *i.i.d.* disturbances).

Parameters	Posterior mode	Posterior st. err.	MLE
μ	1.980	0.363	1.457
α_1	0.6	0	0.6
α_2	0.152	0.132	0.090
θ_1	−0.610	0.175	−0.612
θ_2	0.567	0.147	0.553
$b\left(S^P_t = 1\right)$	1.264	0.185	1.419
$\sigma^2_\xi \left(S^\xi_t = 1 \right)$	0.508	0.073	0.407
$\sigma^2_\xi \left(S^\xi_t = 2 \right)$	1.263	0.169	1.335
$\sigma^2_\zeta \left(S^\zeta_t = 1 \right)$	0.536	0.053	0.546
$\sigma^2_\zeta \left(S^\zeta_t = 2 \right)$	4.097	0.613	4.289
q^ξ_{11}	0.992	0.005	0.985
q^ξ_{22}	0.984	0.009	0.936
q^ζ_{11}	0.994	0.004	0.996
q^ζ_{22}	0.988	0.009	0.985

Table 8: Estimates of the best fitting Markov switching model

References

Albanesi, S., V. V. Chari, and L. J. Christiano (2001): "How Severe is the Time Inconsistency Problem in Monetary Policy?," mimeo, Northwestern University.

——— (2003): "Expectations Traps and Monetary Policy," mimeo, Northwestern University.

Alogoskoufis, G. S. (1992): "Monetary Accomodation, Exchange Rate Regimes and Inflation Persistence," *The Economic Journal*, 102(412), 461–480.

Alogoskoufis, G. S., and R. Smith (1991): "The Phillips Curve, The Persistence of Inflation, and the Lucas Critique: Evidence from Exchange-Rate Regimes," *The American Economic Review*, 81(5), 1254–1275.

Barro, R. J. (1979): "On the Determination of Public Debt," *Journal of Political Economy*, 87, 940–971.

Barro, R. J., and D. B. Gordon (1983): "A Positive Theory of Monetary Policy in a Natural Rate Model," *The Journal of Political Economy*, 91(4), 589–610.

Beck, G. W., and V. Wieland (2002): "Learning and Control in a Changing Economic Enviroment," *Journal of Economic Dynamics and Control*, 26, 1359–1377.

Benati, L. (2003): "Evolving Post-World War II U.K. Economic Performance," mimeo, Bank of England.

Bernanke, B. S., and I. Mihov (1998): "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113, 869–902.

Blanchard, O. J., and S. Fischer (1989): *Lectures on Macroeconomics*. MIT Press, Cambridge, Massachusetts.

Blanchard, O. J., and R. Perotti (2002): "An Empirical Investigation of the Dynamic Effect of Shocks to Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117-4, 1329–1368.

Blanchard, O. J., and J. Simon (2001): "The Long and Large Decline in U.S. Output Volatility," *Brooking Papers on Economic Activity*, 2001-1, 135–174.

Blinder, A. S. (1979): *Economic Policy and the Great Stagflation*. Academic Press, New York.

——— (1998): *Central Banking in Theory and Practice*. MIT Press, Camdridge, Massachusetts.

Boivin, J., and M. P. Giannoni (2002): "Assessing Changes in the Monetary Trasmission Mechanism: A VAR Approach," *Federal Reserve Bank of New York Economic Policy Review*, 8(1), 97–111.

Bullard, J., and S. Eusepi (2003): "Did the Great Inflation Occur Despite Policymaker Committment to a Taylor Rule?," mimeo, Federal Reserve Bank of St. Louis.

Calvo, G. (1983): "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12(3), 383–98.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541–553.

Chari, V. V., L. J. Christiano, and M. Eichenbaum (1998): "Expectations Trap and Discretion," mimeo, Northwestern University.

Cho, I.-K., N. Williams, and T. J. Sargent (2002): "Escaping Nash Inflation," *Review of Economic Studies*, 69, 1–40.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2001): "Nominal Rigidities and the Dynamic Effect of a Shock to Monetary Policy," mimeo, Northwestern University.

Christiano, L. J., and T. Fitzgerald (2003): "Inflation and Monetary Policy in the 20th Century," *Federal Reserve Bank of Chicago Economic Perspectives*, 2003(1), 22–45.

Christiano, L. J., and C. Gust (2000): "The Expectations Trap Hypothesis," *Federal Reserve Bank of Chicago Economic Perspectives*, 2000(2), 21–39.

Chung, H. (1990): "Did Policy Makers Really Believe in the Phillips Curve? An Econometric Test," Ph.D. thesis, University of Minnesota.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115(1), 147–180.

Cogley, T., and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomic Annual*, pp. 331–373, Cambridge, Massachusetts. MIT Press.

——— (2003a): "The Conquest of U.S. Inflation: Learning, Model Uncertainty, and Robustness," mimeo, New York University.

——— (2003b): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," mimeo, New York University.

Debelle, G. (2003): "The Australian Approach to Inflation Targeting," mimeo, Reserve Bank of Australia.

DeLong, B. (1997): "America's Only Peacetime Inflation: The 1970s," in *Reducing Inflation*, ed. by Christina, and D. Romer. University of Chicago Press, Chicago.

Dennis, R. (2001): "The Policy Preferences of the U.S. Federal Reserve," Federal Reserve Bank of San Francisco, Working Paper No. 2001-08.

Durbin, J., and S. J. Koopman (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89(3), 603–616.

Evans, G. W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, NJ.

Fackler, P. L., and M. J. Miranda (2001): "Solving Non-Linear Rational Expectations Models," mimeo, North Carolina State University.

Favara, G., and P. Giordani (2002): "Reconsidering the Role of Money for Output, Prices and Interest Rates," SSE/EFI Working Paper Series in Economics and Finance No. 514.

Favero, C. A., and R. Rovelli (2003): "Macroeconomic Stability and the Preferences of the Fed. A Formal Analysis," *Journal of Money, Credit and Banking*, 35(4), 545–556.

Fernandez-Villaverde, J., and J. Rubio (2002): "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," mimeo, University of Pennsylvania.

Friedman, M. (1968): "The Role of Monetary Policy," *American Economic Review*, 58, 1–17.

Fuhrer, J., and G. Moore (1995a): "Inflation Persistence," *The Quarterly Journal of Economics*, 110(1), 127–159.

Fuhrer, J. C., and G. R. Moore (1995b): "Monetary Policy Trade-Offs and the Correlation Between Nominal Interest Rates and Real Output," *American Economic Review*, 85-1, 219–239.

Gali, J., and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37(4), 1661–1707.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995): *Bayesian Data Analysis*. Chapman and Hall, London.

Gordon, R. J. (1982): "Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment," in *Workers, Jobs and Inflation*, ed. by M. N. Baily. Brookings Institution.

——— (1997): "The Time-Varying NAIRU and its Implications for Economic Policy," *The Journal of Economic Perspectives*, 11(1), 11–32.

——— (1998): "Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU," *Brooking Papers on Economic Activity*, 1998(2), 297–346.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hanson, M. (2003): "Varying Monetary Policy Regimes: A Vector Autoregressive Investigation," mimeo, Wesleyan University.

Ireland, P. N. (1999): "Does the Time-Consistency Problem Explain the Behavior of Inflation in the United States?," *Journal of Monetary Economics*, 44(2), 279–91.

Judd, J. P., and G. D. Rudebusch (1998): "Taylor's Rule and the Fed: 1970-1997," *Federal Reserve Bank of San Francisco Economic Review*, 98(3), 3–16.

Judd, K. L. (1998): *Numerical Methods in Economics*. The MIT Press, Cambridge, MA.

Kim, C.-J., and C. R. Nelson (1999a): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics*, 81, 608–616.

——— (1999b): *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. M.I.T. Press, Cambridge, Massachussetts.

Kim, C.-J., C. R. Nelson, and J. Piger (2001): "The Less Volatile U.S. Economy: A Bayesian Investigation of Breadth, and Potential Explanations," Board of Governors of the Federal Reserve System, International Finance Discussion Paper no. 707.

Kim, S., N. Shephard, and S. Chib (1998): "Sthochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), 361–393.

King, R., and M. W. Watson (1998): "The Solution of Singular Linear Difference Systems Under Rational Expectations," *International Economic Review*, 39(4), 1015–1026.

King, R. G., J. H. Stock, and M. W. Watson (1995): "Temporal Instability of the Unemployment-Inflation Relationship," *Economic Perspectives of the Federal Reserve Bank of Chicago*, pp. 2–12.

King, R. G., and M. W. Watson (1994): "The Postwar US Phillips Curve: A Revisionist Econometric History," *Carnegie-Rochester Conference Series on Public Policy*, 41, 157–219.

Kydland, F. E., and E. C. Prescott (1977): "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85(3), 473–492.

Lansing, K. J. (2002): "Learning About a Shift in Trend Output: Implications for Monetary Policy and Inflation," Federal Reserve Bank of San Francisco Working Paper 00-16.

Leeper, E., and T. Zha (2002): "Modest Policy Interventions," NBER working paper no. 9192.

Leeper, E. M., and J. E. Roush (2003): "Putting 'M' Back in Monetary Policy," mimeo, University of Indiana.

Lubik, T. A., and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *America Economic Review*, 94(1), 190–217.

Mayer, T. (1999): *Monetary Policy and the Great Inflation in the United State: The Federal Reserve and the Failure of Macroeconomic Policy, 1965-1979*. Edward Elgar, Cheltenam, UK.

McConnell, M. M., and G. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's," *American Economic Review*, 90(5), 1464–1476.

Mountford, A., and H. Uhlig (2002): "What are the Effects of Fiscal Policy Shocks?," Tilburg University, Center for Economic Research Discussion Paper 2002-31.

Onatski, A., and N. Williams (2003): "Modeling Model Uncertainty," *Journal of the European Economic Association*, 1(5), 1087–1122.

Orphanides, A. (2000): "The Quest for Prosperity Without Inflation," European Central Bank Working Paper No. 15.

——— (2002): "Monetary Policy Rules and the Great Inflation," *American Economic Review*, 92(2), 115–120.

Orphanides, A., and S. Van Norden (2001): "The Unreliability of Output Gap Estimates in Real Time," CIRANO working paper 2001s-57.

Orphanides, A., and J. C. Williams (2002): "Robust Monetary Policy Rules with Unknown Natural Rates," *Brooking Papers on Economic Activity*, 2002(2), 63–145.

Perotti, R. (2002): "Estimating the Effect of Fiscal Policy in OECD Countries," mimeo, European University Institute.

Phelps, E. S. (1967): "Phillips Curves, Expectations and Optimal Unemployment over Time," *Economica*, 34, 254–281.

Primiceri, G. (2003): "Time Varying Structural Vector Autoregressions and Monetary Policy," mimeo, Princeton University.

Reis, R. (2003): "What is the Natural Rate? Rational Policy Mistakes and Persistent Deviations of Inflation from Target," mimeo, Harvard University.

Rogoff, K. (2003): "Globalization and Global Disinflation," mimeo, International Monetary Fund.

Romer, C., and D. Romer (2002): "The Evolution of Economic Understanding and Postwar Stabilization Policy," in *Rethinking Stabilization Policy*. Federal Reserve Bank of Kansas City.

Romer, C. D., and D. H. Romer (2003): "A New Measure of Monetary Shocks: Derivation and Implications," mimeo, University of California at Berkeley.

Rudebusch, G. D., and L. E. O. Svensson (1999): "Policy Rules for Inflation Targeting," in *Monetary Policy Rules*, ed. by J. B. Taylor, pp. 203–246. University of Chicago Press.

Samuelson, P. A., and R. M. Solow (1960): "Analytic Aspects of Anti-Inflation Policy," *American Economic Review*, 50, 368–379.

Sargent, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press, Princeton, New Jersey.

Schervish, M. J. (1994): *Theory of Statistics*. Springer-Verlag, New York.

Schorfheide, F. (2003): "Learning and Monetary Policy Shifts," mimeo, University of Pennsylvania.

Sims, C. A. (1988): "Projecting Policy Effects with Statistical Models," *Revista de Analisis Economico*, 1988(3), 3–20.

——— (1999): "Drift and Breaks in Monetary Policy," mimeo, Princeton University.

——— (2001a): "Solving Linear Rational Expectations Models," *Journal of Computational Economics*, 20(1-2), 1–20.

——— (2001b): "Stability and Instability in US Monetary Policy Behavior," mimeo, Princeton University.

——— (2002): "Testing Restrictions and Comparing Models," lecture notes, Princeton University.

Sims, C. A., and T. Zha (2004): "Were There Regime Switches in US Monetary Policy?," mimeo, Princeton University.

Smets, F., and R. Wouters (2003): "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Soderlind, P., U. Soderstrom, and A. Vredin (2002): "New-Keynesian Models and Monetary Policy: A Reexamination of the Stylized Facts," Sveriges Riksbank, Working Paper No. 140.

Solow, R. M. (1968): "Recent Controversy on the Theory of Inflation," in *Symposium on Inflation: Its Causes, Consequences and Control*, ed. by S. W. Rousseas, pp. 2–17, New York University.

Staiger, D., J. H. Stock, and M. W. Watson (1997): "The NAIRU, Unemployment and Monetary Policy," *The Journal of Economic Perspectives*, 11(1), 33–49.

——— (2001): "Prices, Wages and the U.S. NAIRU in the 1990s," mimeo, Princeton University.

Stock, J. H., and M. W. Watson (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93(441), 349–358.

——— (2002): "Has the Business Cycle Changed and Why?," NBER working paper no. 9127.

——— (2003a): "Has the Business Cycle Changed? Evidence and Explanations," mimeo, Princeton University.

——— (2003b): "Understanding Changes in International Business Cycle Dynamics," mimeo, Princeton University.

Tambalotti, A. (2003): "Inflation, Productivity and Monetary Policy: From the Great Stagflation to the New Economy," mimeo, Federal Reserve Bank of New York.

Taylor, J. B. (1999): "A Historical Analysis of Monetary Policy Rules," in *Monetary Policy Rules*, ed. by J. B. Taylor, pp. 319–341. University of Chicago Press.

U.S. Office of the President (various years): *Economic Report of the President*. United States Government Printing Office, Washington.

Wieland, V. (2000a): "Learning by Doing and the Value of Optimal Experimentation," *Journal of Economic Dynamics and Control*, 24(4), 501–534.

——— (2000b): "Monetary Policy, Parameter Uncertainty and Optimal Learning," *Journal of Monetary Economics*, 46(1), 199–228.

Williams, N. (2003): "Escape Dynamics in Learning Models," mimeo, Princeton University.

Woodford, M. (2003): "Optimal Interest-Rate Smoothing," *The Review of Economic Studies*, 70(4), 861–886.

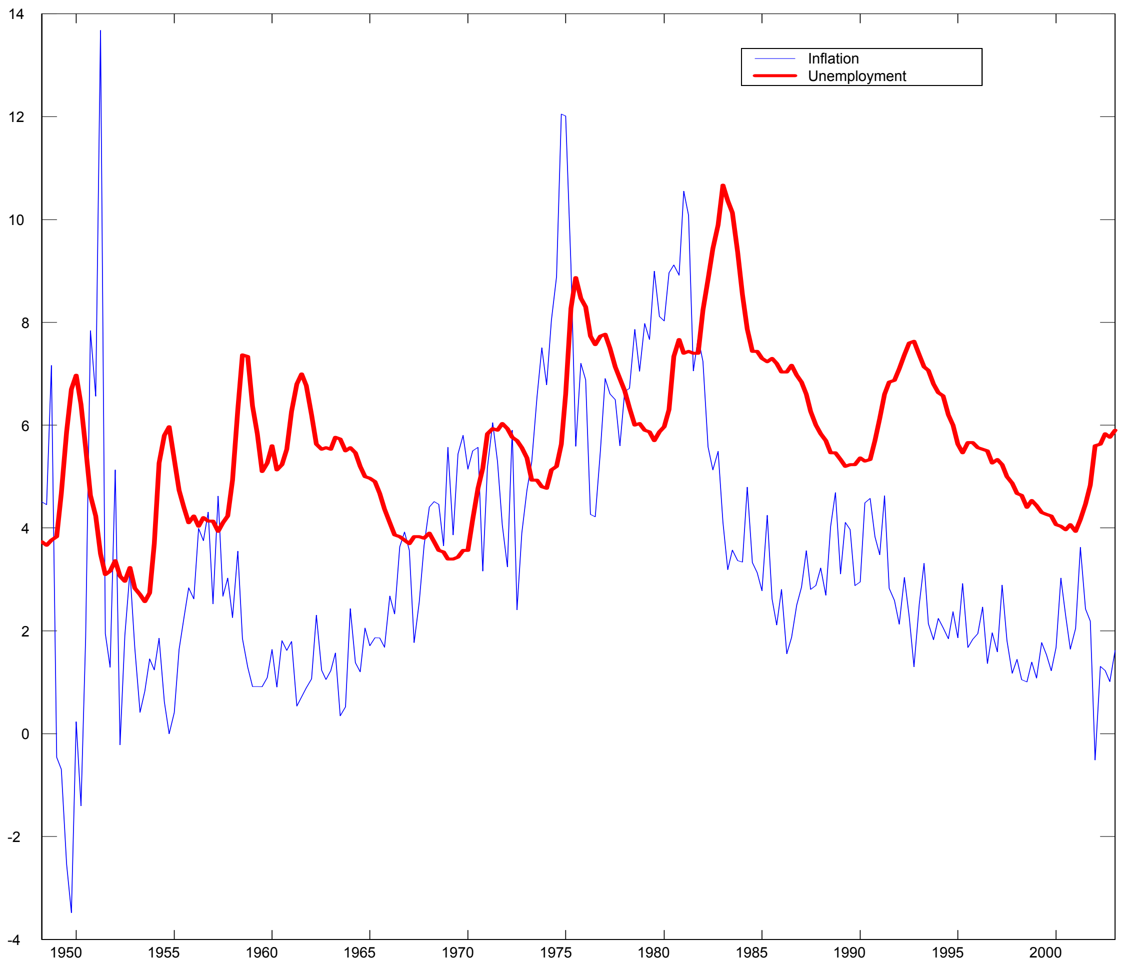

Figure 1: US inflation and unemployment.

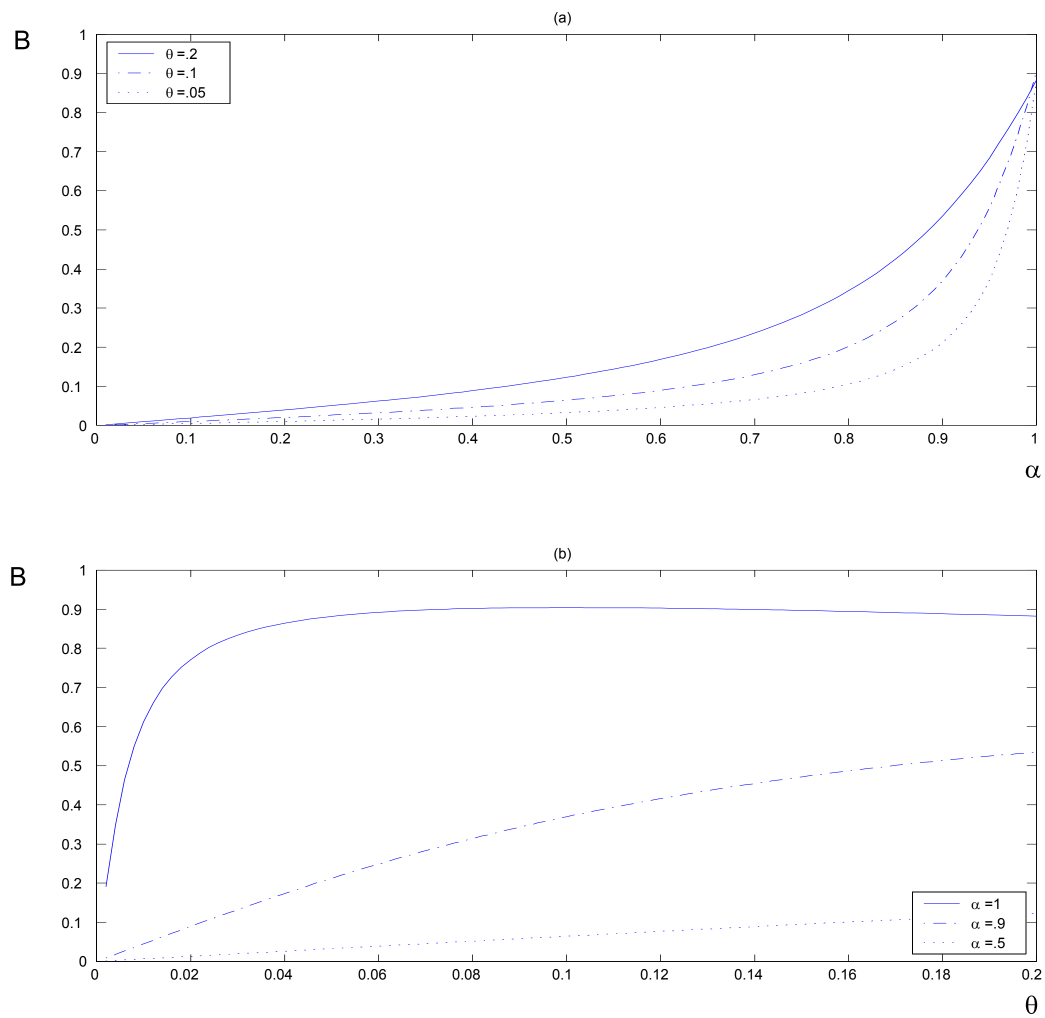

Figure 2: Strength of the policy reaction to inflation as a function of (a) estimated persistence of inflation in the Phillips curve and (b) estimated slope of the Phillips curve.

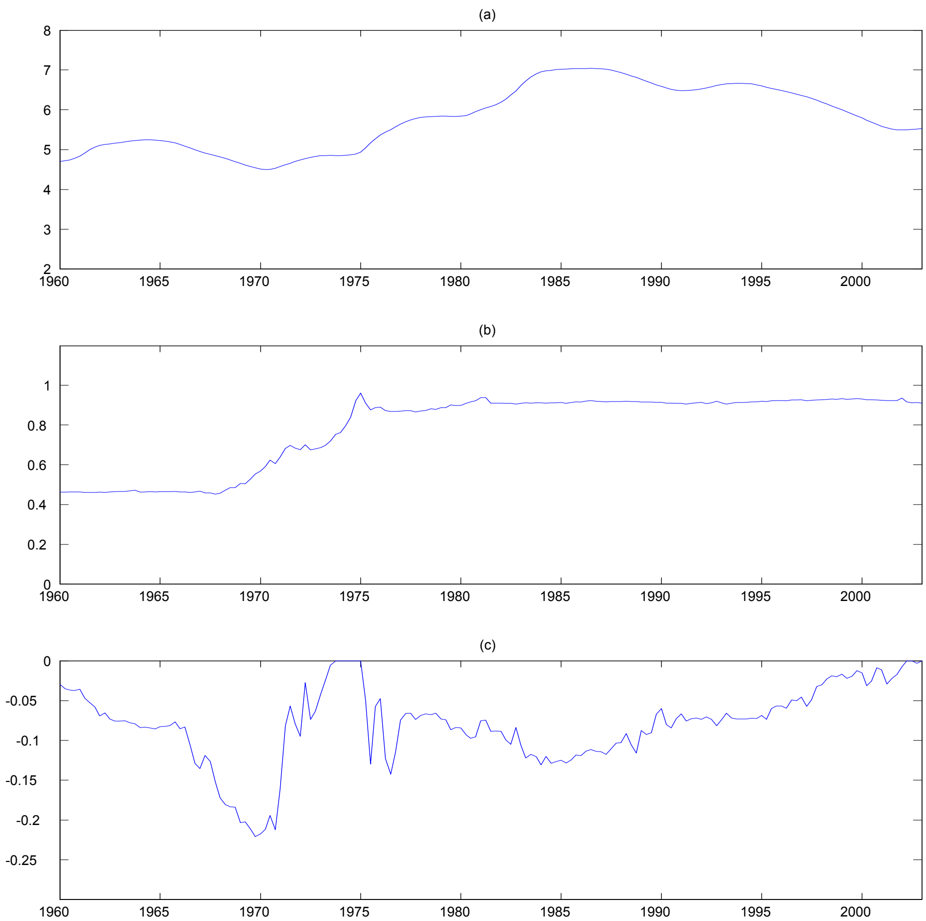

Figure 3: Evolution of policymakers' beliefs about: (a) the non-accelerating inflation rate of unemployment; (b) the persistence of inflation in the Phillips curve; (c) the slope of the Phillips curve.

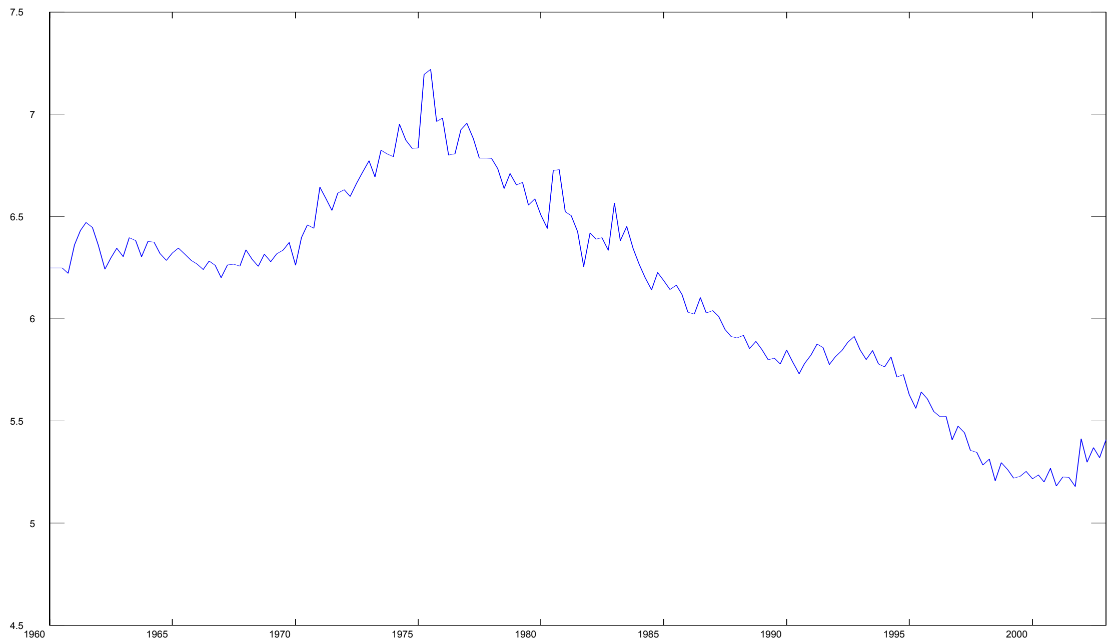

Figure 4: Smoothed estimates of the non-accelerating inflation rate of unemployment.

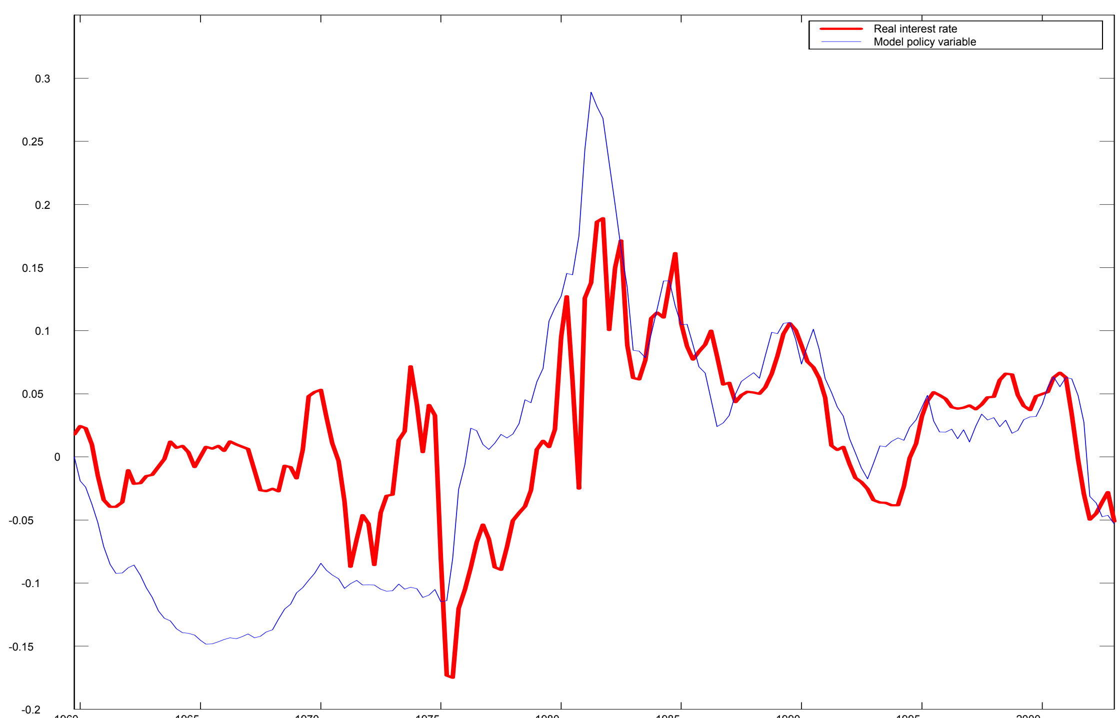

Figure 5: Models' policy variable and real rate of interest (rescaled).

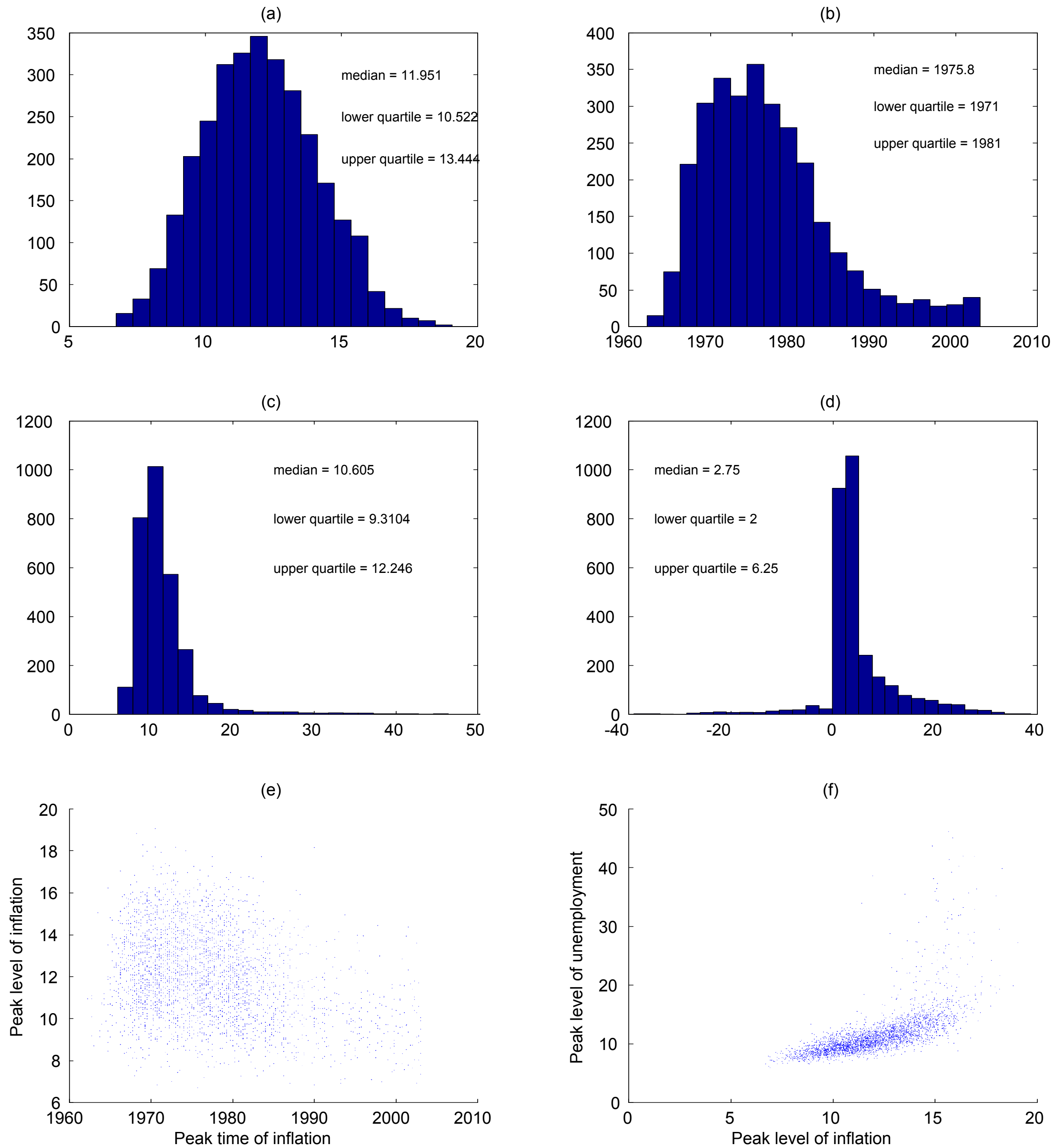

Figure 6: Simulation results for the benchmark model. Empirical distribution of: (a) $\max(\pi)$, (b) t^*_π, (c) $\max(u)$ and (d) $t^*_u - t^*_\pi$. Scatter plot of the relation between (e) t^*_π and $\max(\pi)$ and (f) $\max(\pi)$ and $\max(u)$. ($\max(\pi)$ and t^*_π stand respectively for the peak level and the peak time of inflation, while $\max(u)$ and t^*_u stand for the peak level and the peak time of unemployment).

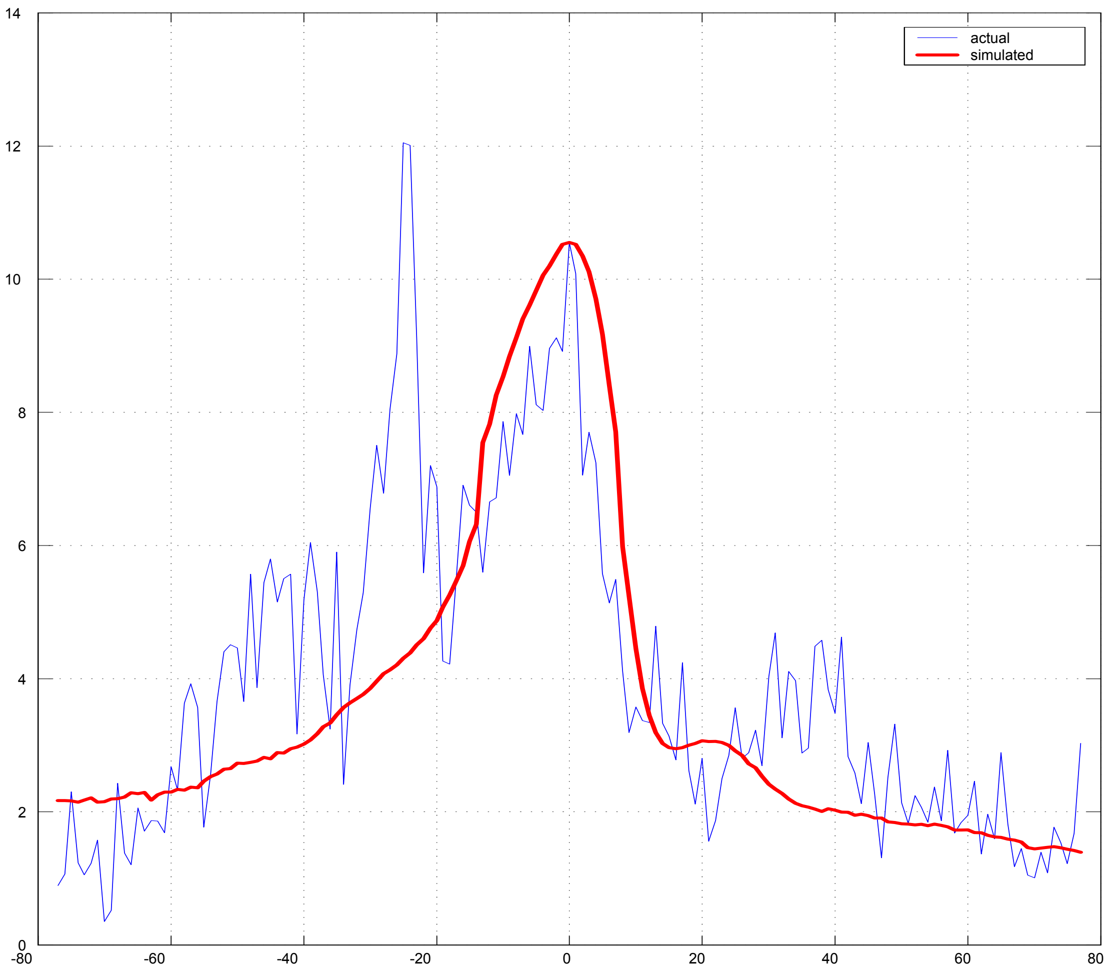

Figure 7: Actual data and average of the simulated inflation paths around the peak time (time expressed in quarters on the horizontal axis).

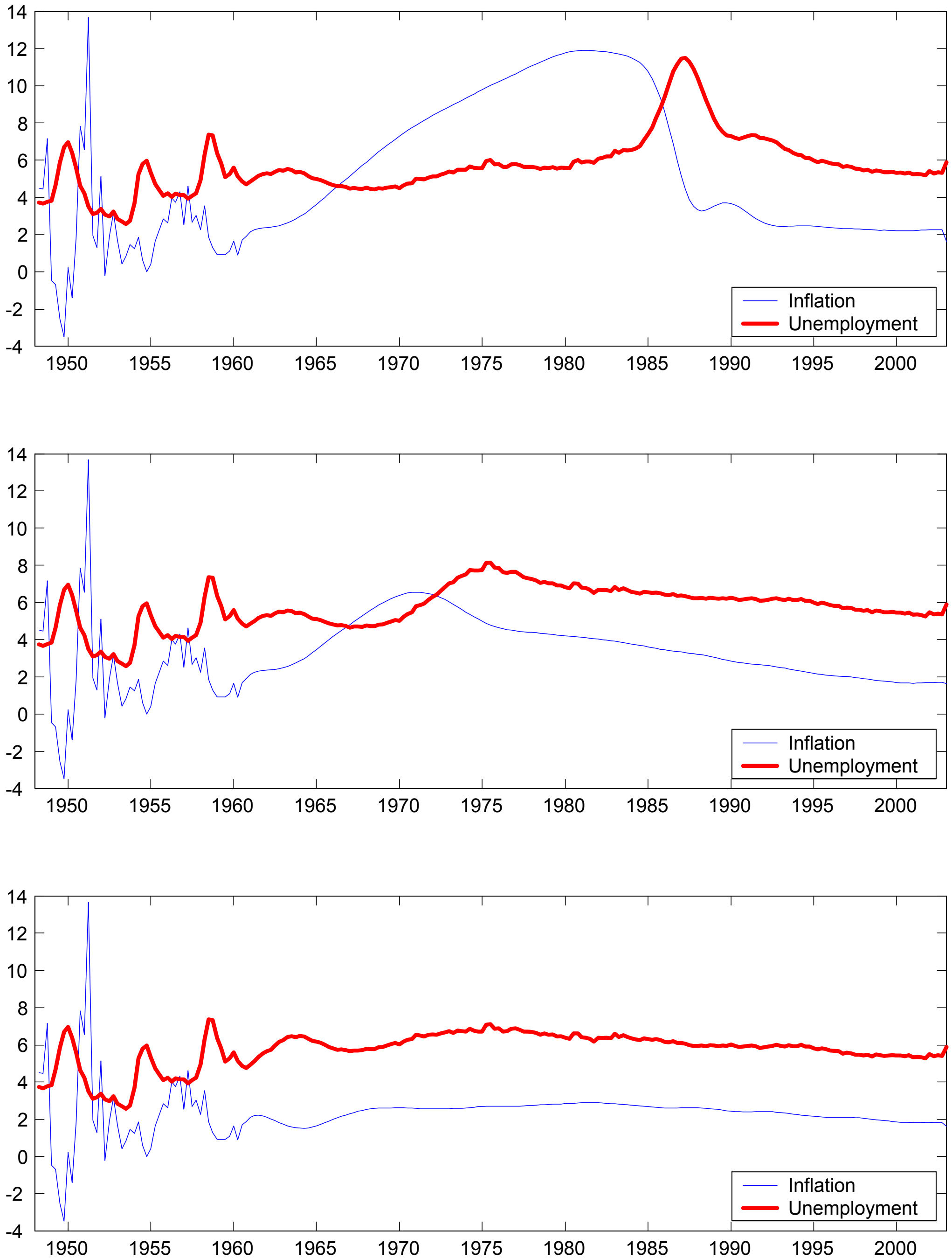

Figure 8: Simulation of the inflation and unemployment behavior under a scenario of low volatility of the exogenous disturbances. (a) Simulation for the baseline constant gain learning model; (b) counterfactual simulation under the assumption that policymakers know the slope of the Phillips curve; (c) counterfactual simulation under the assumption that policymakers know all the parameters of the Phillips curve except for the non-accelerating inflation rate of unemployment.

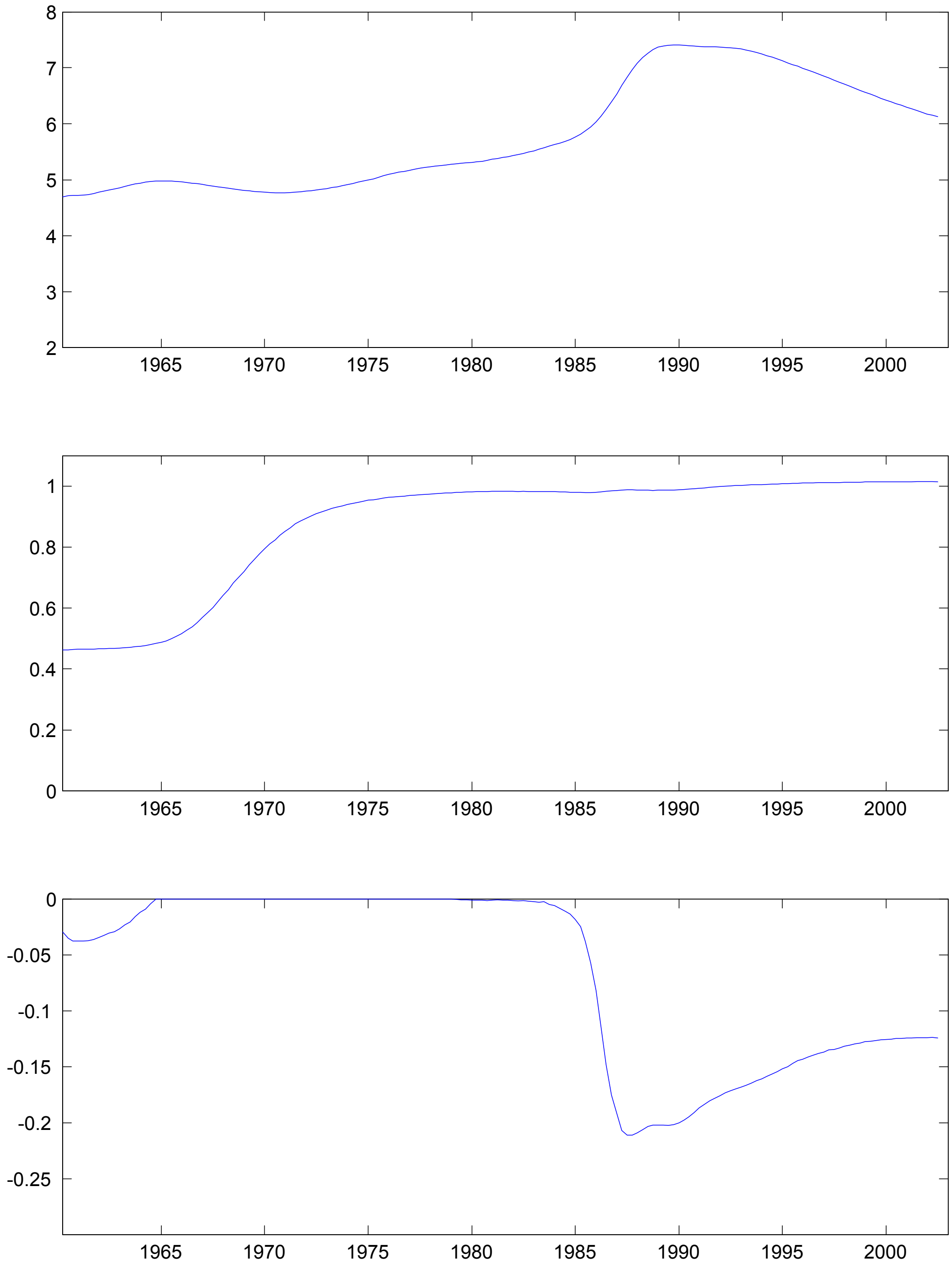

Figure 9: Simulation of the evolution of policymakers' beliefs under a scenario of low volatility of the exogenous disturbances: (a) the non-accelerating inflation rate of unemployment; (b) the persistence of inflation in the Phillips curve; (c) the slope of the Phillips curve.

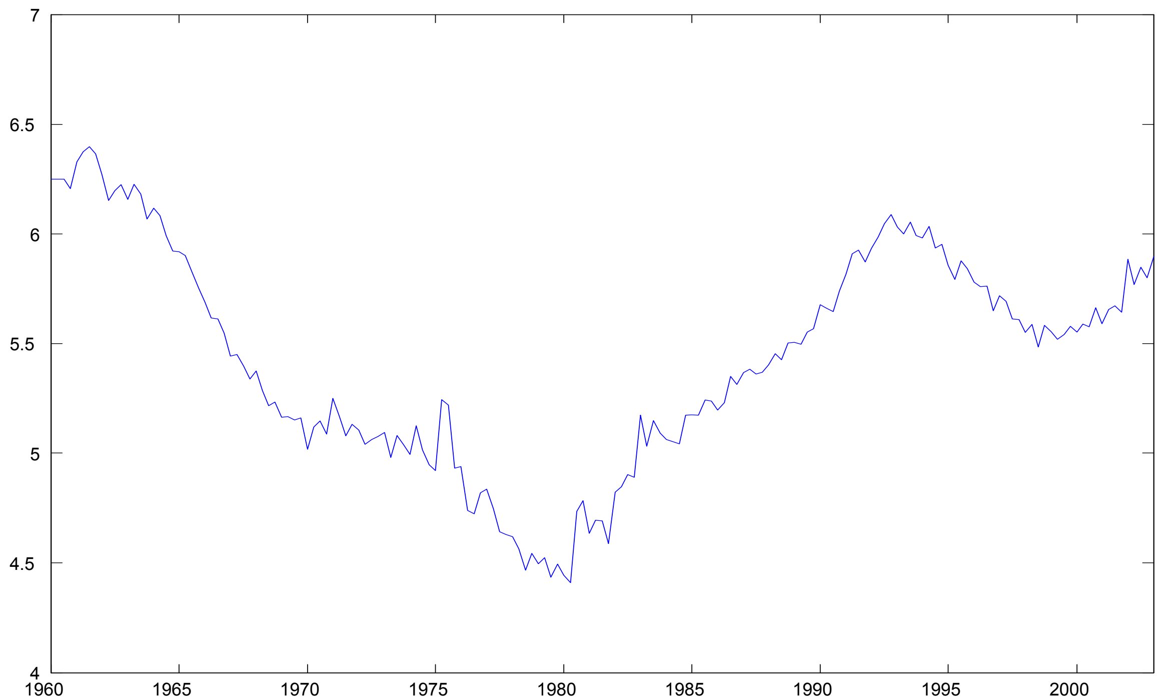

Figure 10: Smoothed estimates of the non-accelerating inflation rate of unmployment obtained from a model in which policymakers know the true values of the persistence of inflation and the slope of the Phillips curve.

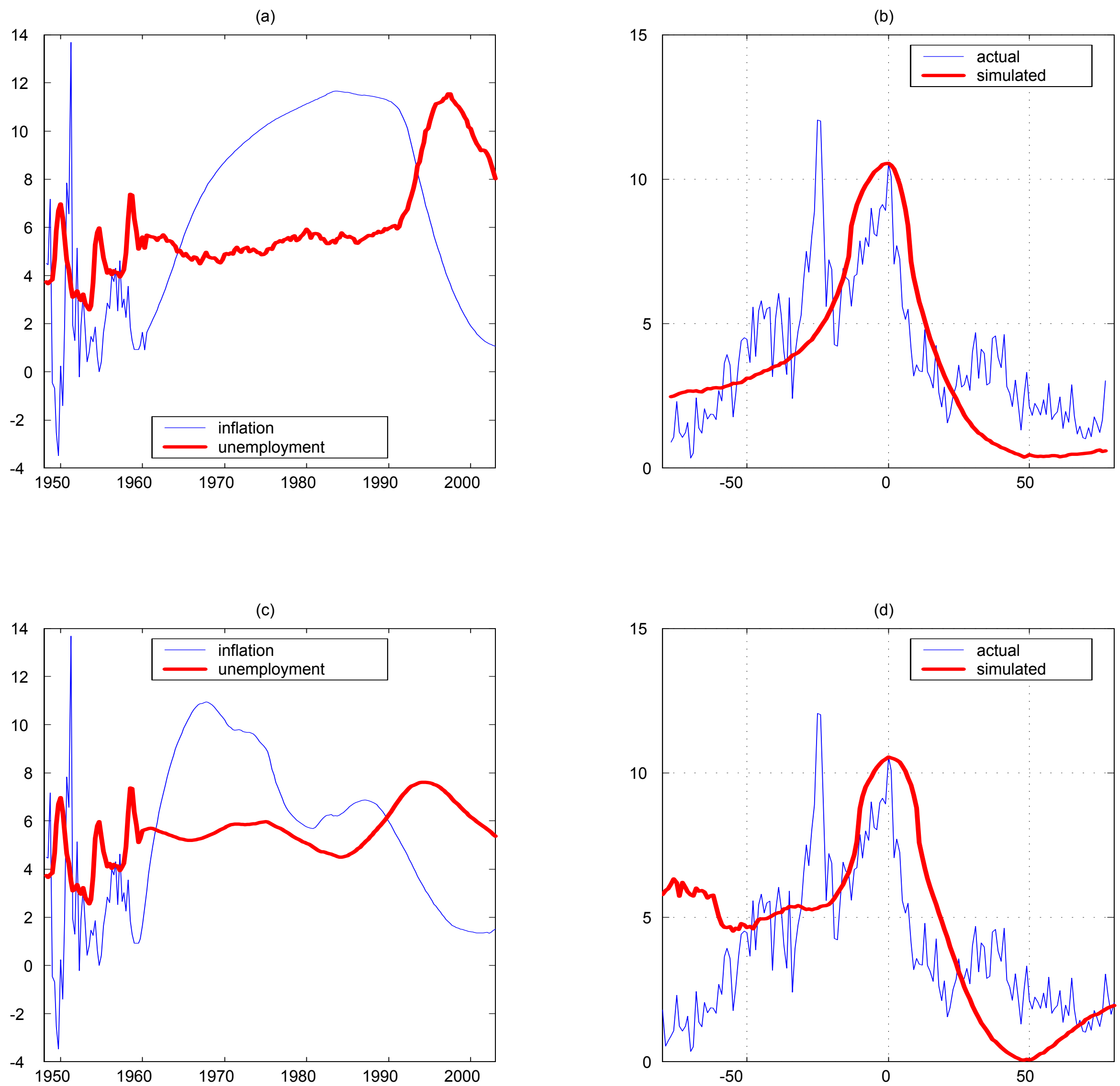

Figure 11: Simulation of inflation and unemployment under a scenario of low volatility of the exogenous disturbances in (a) the partially rational agents model of section 5.1 and (c) the fully rational agents model of section 5.2. Average of all simulated inflation paths around the peak under realistic volatility of the exogenous disturbances in (b) the partially rational agents model and (d) the fully rational agents model.

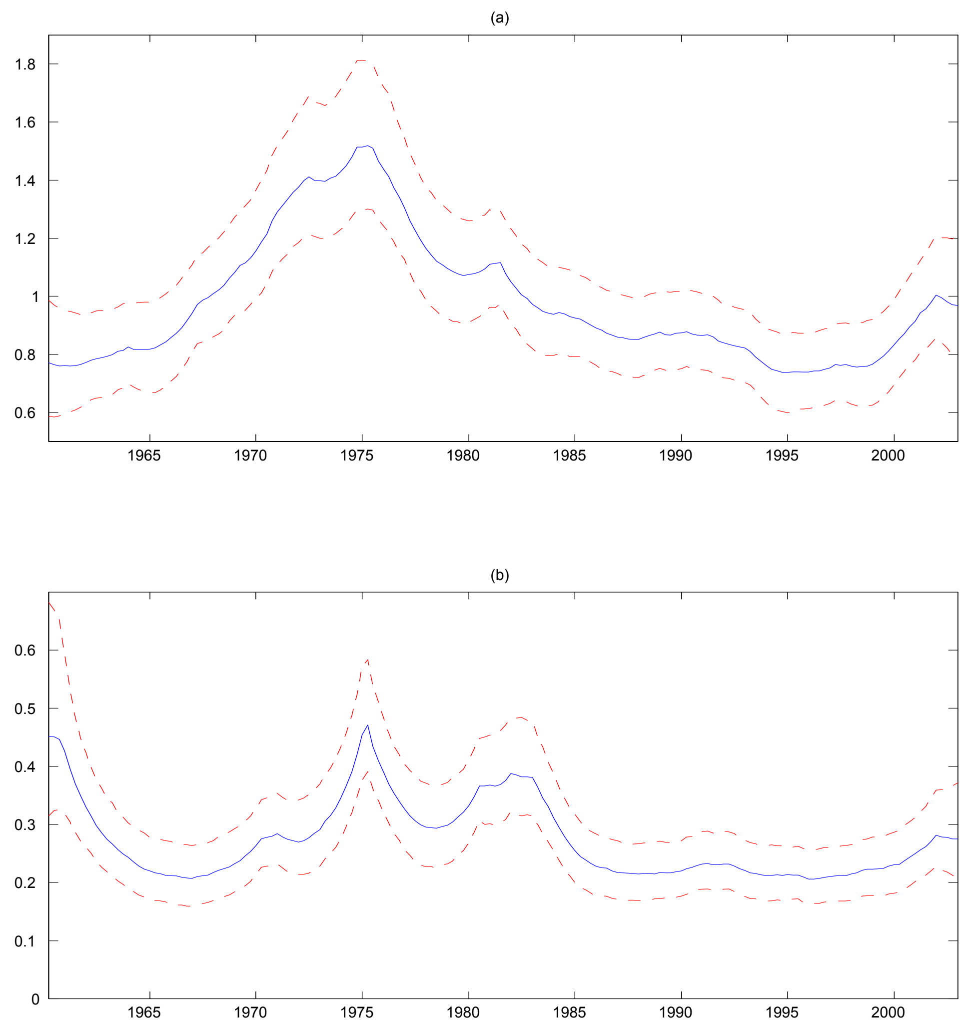

Figure 12: Posterior median and 68% error bands for the time varying standard deviations of (a) shocks to the Phillips curve and (b) shocks to the aggregate demand.

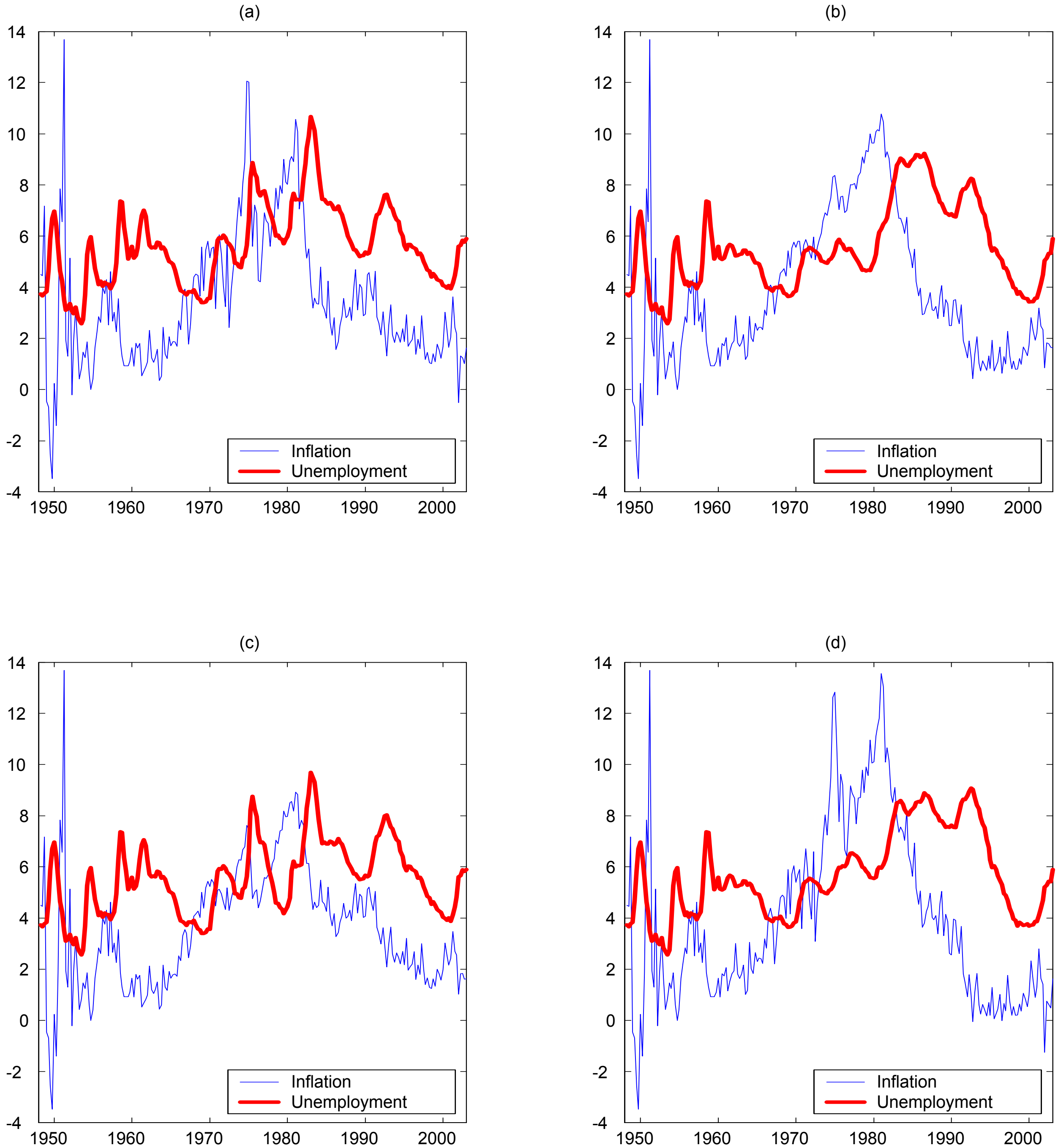

Figure 13: (a) Actual data for inflation and unemployment and counterfactual simulations when (b) the standard deviations of the shocks to the Phillips curve and the aggregate demand are fixed to their 1995 value; (c) only the standard deviation of the shocks to the Phillips curve is fixed to its 1995 value; (d) only the standard deviation of the shocks to the aggrgate demand is fixed to its 1995 value.

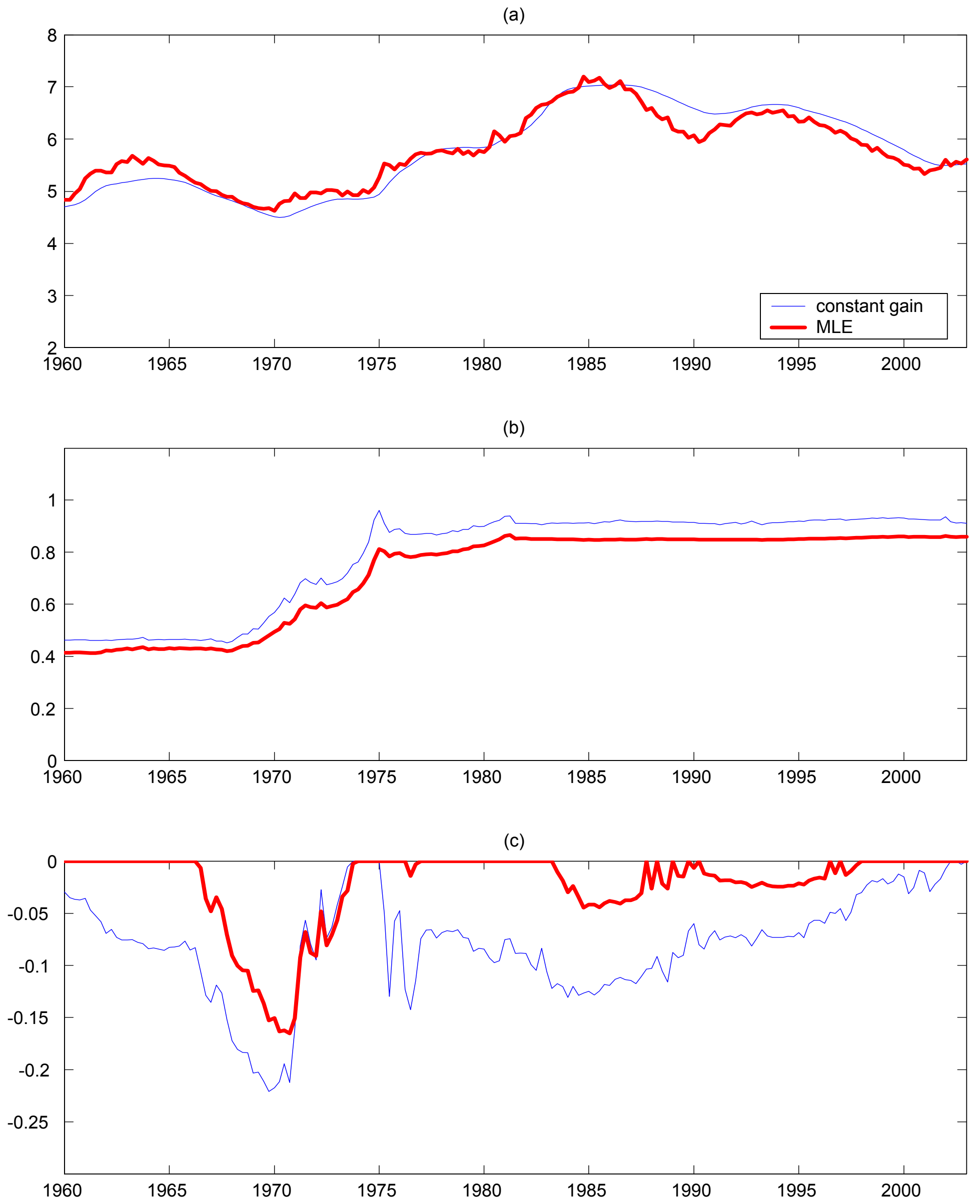

Figure 14: Evolution of policymakers' beliefs about: (a) the non-accelerating inflation rate of unemployment; (b) the persistence of inflation in the Phillips curve; (c) the slope of the Phillips curve. Results for the baseline constant gain learning and the maximum likelihood learning of section 6.2.

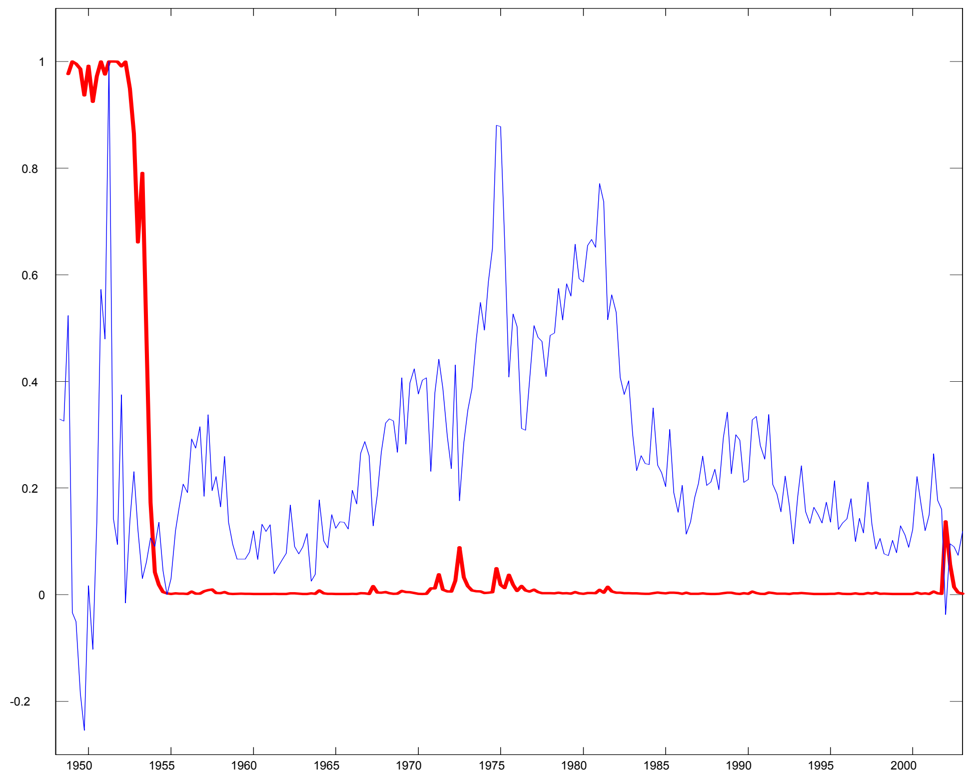

Figure 15: Inflation rate and posterior probability of the high volatility regime of the *noise* component. See section 7 for details.

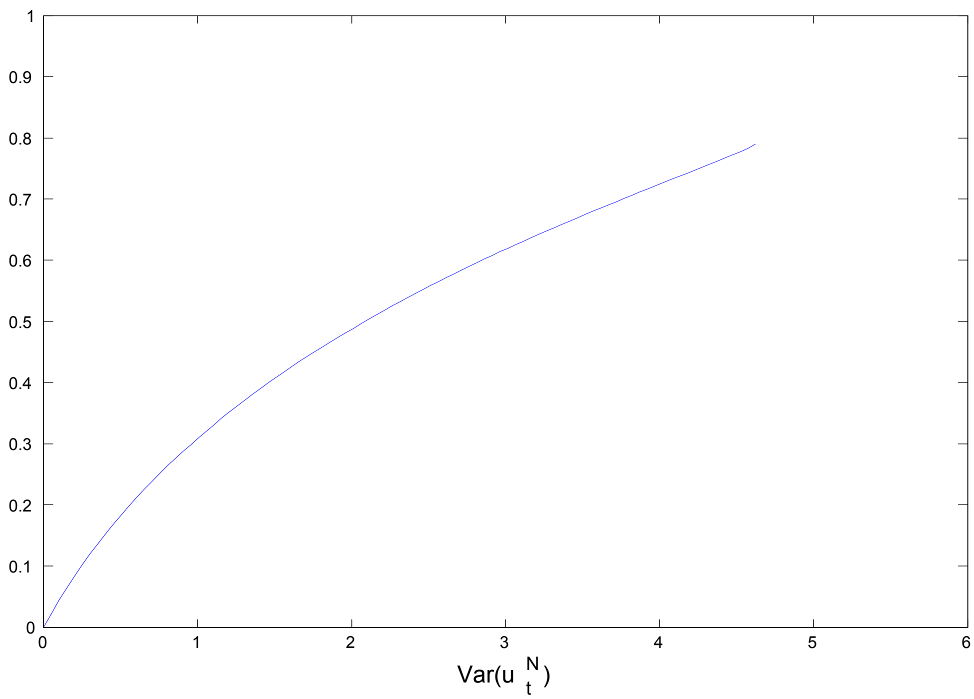

Figure 16: Euclidean norm of the distance between the true value of the coefficients and the value of policymakers beliefs in a self-confirming equilibrium (as a function of the variance of the non-inflationary rate of unemployment).