

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 962

December 2008

Uncertainty Over Models and Data: The Rise and Fall of American Inflation

Seth Pruitt

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgement that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/. This paper can be downloaded without charge from Social Science Research Network electronic library at www.ssrn.com

Uncertainty over Models and Data: The Rise and Fall of American Inflation*

Seth Pruitt[†]
Federal Reserve
Board of Governors

December 2008

Abstract

An economic agent who is uncertain of her model updates her beliefs in response to the data. The updating is sensitive to measurement error which, in many cases of macroeconomic interest, is apparent from the process of data revision. I make this point through simple illustrations and then analyze a recent model of the Federal Reserve's role in U.S. inflation. The existing model succeeds at fitting inflation to optimal policy, but fails to link inflation to the economic trade-off at the heart of the story. I modify the model to account for data uncertainty and find that doing so ameliorates the existing problems. This suggests that the Fed's model uncertainty is largely overestimated by ignoring data uncertainty. Consequently, now there is an explanation for the rise and fall in inflation: the concurrent rise and fall in the perceived Philips curve trade-off.

Keywords: data uncertainty, data revisions, real time data, optimal control, parameter uncertainty, learning, extended kalman filter, markov-chain monte carlo

JEL codes: E01, E58

*First version: February 2007. I thank Jim Hamilton for his insightful advice. Thanks also to Alina Barnett, Gray Calhoun, Martin Ellison, Margie Flavin, Nir Jaimovich, Garey Ramey, Valerie Ramey, Chris Sims, Allan Timmermann, Harald Uhlig, Chris Wignall, and Tao Zha for helpful discussions; to seminar participants at UCSD, Georgetown, HEC Montreal, Texas A&M, Penn State, UCSC, Cambridge University Learning and Macroeconomic Policy Conference, Philadelphia Fed, Kansas City Fed, Atlanta Fed, and the Federal Reserve Board; and to Tao Zha for kindly providing computer code and data. This paper was previously circulated as part of the deprecated "When Data Revisions Matter to Macroeconomics."

[†]Division of International Finance, seth.j.pruitt@frb.gov. This paper reflects the views and opinions of the author and does not reflect on the views or opinions of the Federal Reserve Board, the Federal Reserve System, or their respective staffs.

1 Introduction

A great deal of research has gone towards identifying the causes of the large swings in inflation in the United States between 1970 and 1985. One strand of literature advances the view that evolving government responses had an important role in these events. Clarida, Gali, and Gertler (2000) provides evidence of different U.S. monetary policy responses over different parts of the postwar era. Boivin (2006) elaborates on this finding using a time-varying Taylor rule estimated on real-time data, as Primiceri (2005) does in a DSGE model using revised data, to characterize the response without giving explanation for its evolution. Romer and Romer (1990) and Owyang and Ramey (2004) suggest that the changing response might be explained by changing Fed objectives. On the other hand, Sargent (1999) suggests that Federal Reserve beliefs alone, evolving in response to data, can explain the rise and fall of inflation.¹

This last explanation relies on the idea that agents learn about their economic environment. In such a model, agents' prediction errors update their beliefs, represented as the parameters to their model of their world. Obviously, these prediction errors are based on the observed data. Orphanides (2001) notes that researchers should carefully attribute to agents the data actually observable at the time of decision-making.² Because some data is revised, real-time data available to the agent may be different than the revised data available to the researcher. The point of my paper is that agents' data uncertainty, resulting from their knowledge that real-time data is revised, affects the evolution of their beliefs because they know it can affect their prediction errors. Therefore researchers miscalculate the volatility of these beliefs, with a tendency towards overestimation, by ignoring observable data uncertainty.

I explain this point through simple frameworks that have two main ingredients. First, the agent has model uncertainty represented as a parameter shock or error that explicitly allows the agent to change their beliefs in response to observed data. Second, the agent has data uncertainty represented as an error that causes data to be mismeasured. The conclusion from this analysis is that the researcher's estimate of the agent's model uncertainty is biased by ignoring existing data uncertainty, and that this bias is positive in many simple cases and more complicated cases subject to empirically-relevant restrictions.

¹Orphanides and Williams (2006) and Sims (2006) also argue that policy is sensitive to model uncertainty, and Nason and Smith (forthcoming) finds instability in the Philips curve relationship since the 1950s.

²Runkle (1998) made a similar point around the same time.

Then I take an example from the fore mentioned literature in order to understand the effect of this bias in practice. As application I use the framework in Sargent, Williams, and Zha (2006) which follows the main idea of Sargent (1999). The Federal Reserve optimally controls inflation in light of unchanging unemployment and inflation targets, however the Fed is uncertain of its economic model and thus lets its beliefs evolve in response to new data.³ Hence, Sargent, Williams, and Zha (2006) aim to explain the great American inflation as optimal policy given changing estimates of the Philips curve. However, the model suffers from three key problems. One, the Fed's unemployment rate forecasts, the basis for setting inflation, are very inaccurate, much more so than the Greenbook forecasts they should mimic. Second, the Fed's estimated model uncertainty is very large, which undermines the plausibility that the Fed believed in its estimated Philips curve enough to use it as the basis for policy. Third, the model explains the rise in inflation between 1973 and 1975, but does not give a good reason for the drastic fall in inflation between 1980 and 1984.

I adapt the Fed's estimation problem to account for the observable fact that macroeconomic data is revised.⁴ By accounting for data uncertainty, on which we have actual evidence, the model explains inflation without the fore mentioned problems. The Fed is confident of its model and makes accurate unemployment forecasts that resemble Greenbook forecasts. Importantly, the model now predicts a sharp drop in the Philips trade-off between 1980 and 1984 leading to the concomitant drastic drop of inflation.

The economic difference from Sargent, Williams, and Zha (2006) is that here the Fed is sluggish to change its beliefs in response to real-time data because this data might be revised. Since at least as far back as Zellner (1958) – who admonished readers to be “careful” with “provisional” data – the reality of data uncertainty has been clearly recognized.⁵ What has been less clear is the impact this data uncertainty may hold for the purposes of modeling. My paper shows that data uncertainty can have a large effect in models where agents change their beliefs in response to incoming data.

In addition to fore mentioned studies on the great inflation, my paper joins literature highlighting the impact of real-time data on economic behavior. This is made possible by the pioneering work

³Model uncertainty in this case is represented by the variance of shocks to time-varying parameters.

⁴Note that my model reduces to Sargent, Williams, and Zha (2006)'s model when one assumes data is seen without error and, therefore, later vintages of data are identical to its first report.

⁵In fact, data uncertainty was recognized by Burns and Mitchell, who revised their business cycle indicators as data revisions came in, and considered many macroeconomic variables for that reason.

of Croushore and Stark (2001) on the Philadelphia Fed’s Real-Time Data Set for Macroeconomists, and the St. Louis Fed’s ensuing Archival Federal Reserve Economic Data (ALFRED) collection. In early work, Oh and Waldman (1990) used data revisions to identify the effects of real-time macro announcements on future economic activity, highlighting how data mismeasurement itself can influence agents’ behavior. Ghysels, Swanson, and Callan (2002) extended this idea by using multiple data vintages to fit policy rules, some of which are adaptively estimated, forecasting the Fed Funds rate. These papers suggest that agents’ behavior is connected to the real-time data they actually saw when making decisions. Within a DSGE framework, Aruoba (2004) analyzed the welfare consequences of a one-period signal extraction problem motivated by data uncertainty. Many recent papers by central bank researchers are evidence that macroeconomic data revisions remain significant to policy decisions even today; see Cunningham, Jeffery, Kapetanios, and Labhard (2007) and references therein.

Explicitly modeling data uncertainty is similar in spirit to the macroeconomic learning literature where the main idea is that agents are themselves econometricians. Whereas much of that literature focuses on the problem of learning parameters – for instance see Evans and Honkapohja (2001) or Orphanides and Williams (2006) – I focus attention to the problem of learning the true economic state measured by imperfect real-time data. In addressing the problematic implications in Sargent, Williams, and Zha (2006), I join concurrent work in Carboni and Ellison (2007) and Carboni and Ellison (2008). The first paper changes the Fed’s objective to include the goal of accurate unemployment forecasts, resulting in a new policy rule sensitive to the accuracy of parameter estimates; the second paper changes the unemployment forecast estimation to target Greenbook forecasts. In particular, the second paper achieves success at addressing the problems regarding unemployment rate forecasts and large model uncertainty.⁶ However, neither paper addresses the missing explanation for the rise *and fall* of inflation.

The paper is organized as follows. Section 2 provides a simple explanation of the impact of data uncertainty when there exists model uncertainty. Section 3 evaluates the impact of data uncertainty in practice by applying it to the Sargent, Williams, and Zha (2006) framework. Section 4 presents the estimation results of both the model without data uncertainty and the model including data uncertainty; it is apparent that the latter remedies the problems of the former, and Section 5

⁶In contrast to Carboni and Ellison (2008), the model in this paper produces unemployment forecasts that are statistically indistinguishable from the Greenbook forecasts without using those Greenbook forecasts as data.

explains why. Section 6 concludes.

2 Why Data Uncertainty Matters

Suppose an agent forms a forecast of an economic quantity y . The agent's model maps observed data and a parameter vector into this prediction. *Model uncertainty* is the situation where past predictions and realized data, at some point in time and through the agent's economic model, might change the parameter vector going forward. *Data uncertainty* is the situation where data is measured with error perhaps, but not necessarily, observed after the fact. The main question here is, if the agent is uncertain both of the model and the data, what is the impact of the researcher ignoring data uncertainty?

To answer this, I consider a few instances of prediction error decomposition, which is the machinery through which agents change their models. The intuition is straightforward that when the agent's prediction is correct there is nothing to change about her beliefs.⁷ When such is the case, the agent has nothing to learn since her model's performance cannot be improved.

When the prediction error is nonzero, the agent has incentive to evaluate her model and learn from the error. At this point, the agent needs to understand *why* the prediction error is nonzero: is it due to unpredictable additive shocks, is it due to mismeasured data, is it due to the model (parameter)? We will see that data uncertainty is not negligible when the economic agent is uncertain of her own model.

In the following, ϵ_i is a shock on the i parameter, ε_j is an error on the j data element, and it is assumed that

$$\mathbb{E}(\epsilon_i) = \mathbb{E}(\varepsilon_j) = 0, \text{Var}(\epsilon_i) = \sigma_i^2 > 0, \text{Var}(\varepsilon_j) = \sigma_j^2 > 0, \mathbb{E}(\epsilon_i \varepsilon_j) = 0 \quad \forall i, j$$

I assume that the data and parameter predictions are almost surely nonzero. The parameter shock can be interpreted as actual variation in the parameters, for instance considered by Cooley and Prescott (1976), Primiceri (2005), and Sargent, Williams, and Zha (2006). Alternatively, the shock can be interpreted as a model error allowed for in each period by the agent, just as an empirical model used by the researcher includes an error. I will consider linear predictors calculated under mean squared error loss.

⁷This follows from the assumptions: the agent has an economic loss function, the agent's prediction is rational with respect to that loss function, and the loss function is minimized when the prediction error is zero. See Elliott, Komunjer, and Timmermann (2005) for more extensive discussion.

One parameter on mismeasured data Suppose that the agent believes that

$$y = (b + \epsilon_b)(x + \epsilon_x)$$

We call b the parameter prediction and x the data prediction. The agent's prediction of y is

$$\hat{y} = bx$$

and her forecast error ($y - \hat{y}$) is

$$b\epsilon_x + x\epsilon_b + \epsilon_b\epsilon_x$$

When the forecast error is zero the agent trivially decomposes this to imply that both the parameter shock and the data error are zero because their unconditional means are zero and there is no more than this information. When the forecast error is nonzero, the decomposition is nontrivial but simple. The prediction of each shock's contribution depends on the size of the forecast error and the average size of the shocks (supposed by the agent). This decomposition is obtained from the forecast error by noting

$$\text{Var}(y - \hat{y}) = b^2\sigma_x^2 + x^2\sigma_b^2 \implies \left\{ \begin{array}{l} \hat{\epsilon}_b = \frac{b^2\sigma_b^2}{b^2\sigma_b^2 + x^2\sigma_x^2}(y - \hat{y}) \\ \hat{\epsilon}_x = \frac{x^2\sigma_x^2}{b^2\sigma_b^2 + x^2\sigma_x^2}(y - \hat{y}) \end{array} \right\}$$

which are well-known linear prediction equations.

Suppose the researcher thinks that $\epsilon_x = 0$ almost surely. This implies that $\sigma_x^2 = 0$ and therefore the researcher deduces from the linear prediction equations an estimator $\tilde{\sigma}_b^2$ of how large on average the agent regards the parameter shock to be

$$\tilde{\sigma}_b^2 = \frac{\text{Var}(y - \hat{y})}{x^2}$$

Said another way, the researcher would use the LHS of the above equation to uncover the σ_b^2 supposedly believed by the agent.

However, since the agent holds data uncertainty, this estimator is actually

$$\tilde{\sigma}_b^2 = \sigma_b^2 + \frac{b^2}{x^2}\sigma_x^2$$

which means that

$$\text{Bias}(\tilde{\sigma}_b^2) = \frac{b^2}{x^2}\sigma_x^2 \tag{2.1}$$

Therefore the researcher's estimate of the agent's σ_b , the shock size unit, is biased upwards. The bias depends on the researcher's data and parameter prediction and the actual variance of the data error.

Two parameters, one on mismeasured data Suppose that the agent believes that

$$y = (b + \epsilon_b)(x + \epsilon_x) + (d + \epsilon_d)$$

We call d a parameter prediction.⁸ The agent's prediction of y is again linear and the forecast error follows naturally.

Again, suppose the researcher thinks that $\epsilon_x = 0$ almost surely. Let us suppose $k\sigma_b = \sigma_d$, $k > 0$ so σ_b is the shock size unit being measured. The researcher then calculates

$$\tilde{\sigma}_b^2 = \frac{\text{Var}(y - \hat{y})}{x^2 + k^2 + 2xk\rho_{bd}}$$

where ρ_{bd} is the correlation between ϵ_b and ϵ_d perceived by the researcher.

However, actually

$$\text{Bias}(\tilde{\sigma}_b^2) = \frac{b^2\sigma_x^2}{x^2 + k^2 + 2xk\rho_{bd}} \quad (2.2)$$

The numerator is almost surely positive; so is the denominator, as the following shows

To show:

$$x^2 + k^2 + 2xk\rho_{bd} > 0$$

Proof: Suppose the contrary. Then $x^2 + k^2 < -2xk\rho_{bd}$ and note both sides are positive. Hence $\frac{x^2}{k^2} + 2 + \frac{k^2}{x^2} < 4\rho_{bd}^2$. WLOG, let $x^2/k^2 = 1 + \delta$ for $\delta \geq 0$. Then $1 + \frac{\delta^2}{1+\delta} < \rho_{bd}^2$. But $\rho_{bd} \in [-1, 1]$. ■

Therefore the researcher's estimate of the average size of the agent's parameter shocks is biased upwards.

One parameter on mismeasured data, mismeasured target Suppose that the agent believes that

$$y - \epsilon_y = (b + \epsilon_b)(x + \epsilon_x)$$

The agent's prediction of y and forecast error naturally follows. If the researcher thinks $\epsilon_y = \epsilon_x = 0$ almost surely, then

$$\text{Bias}(\tilde{\sigma}_b^2) = \frac{b^2\sigma_x^2 + \sigma_y^2 + 2b\rho_{xy}\sigma_x\sigma_y}{x^2} \quad (2.3)$$

⁸We could instead label d/x_0 a parameter prediction, ϵ_d/x_0 a parameter shock, and x_0 be some data measured without error that is multiplied by $(d/x_0 + \epsilon_d/x_0)$, and the idea here would be the same. Likewise if $d = 0$ and ϵ_d is simply an additive error term in which case it might be natural to assume $\rho_{bd} = 0$.

where ρ_{xy} is the true correlation between ε_x and ε_y . The RHS is almost surely positive. The proof is the same as above, noting that one can write $\sigma_y = k\sigma_x$ for some $k > 0$. Therefore the researcher's estimate of the shock size unit is biased upwards.

Two parameters on mismeasured data

$$y = (b + \epsilon_b)(x + \varepsilon_x) + (c + \epsilon_x)(z + \varepsilon_z)$$

Let $k\sigma_b = \sigma_c$ and $m\sigma_x = \sigma_z$ for $k, m > 0$ so σ_b is the shock size unit being measured. Then

$$\text{Bias}(\tilde{\sigma}_b^2) = \sigma_x^2 \frac{b^2 + c^2 m^2 + 2bcm\rho_{xz}}{x^2 + k^2 z^2 + 2kxz\rho_{bc}} \quad (2.4)$$

This bias is positive, following the proof above for both the numerator and denominator.

Three parameters, two on mismeasured data, mismeasured target Suppose that the agent believes that

$$y - \varepsilon_y = (b + \epsilon_b)(x + \varepsilon_x) + (c + \epsilon_x)(z + \varepsilon_z) + (d + \epsilon_d)$$

The agent's prediction of y and forecast error naturally follows. For the researcher who thinks $\varepsilon_y = \varepsilon_x = \varepsilon_z = 0$ almost surely, let $k\sigma_b = \sigma_c$, $m\sigma_b = \sigma_d$, $n\sigma_x = \sigma_z$, and $p\sigma_x = \sigma_y$ for $k, m, n, p > 0$, so σ_b is the shock size unit being measured. Then

$$\text{Bias}(\tilde{\sigma}_b^2) = \sigma_x^2 \frac{b^2 + c^2 n^2 + p^2 + 2bcn\rho_{xz} + 2bp\rho_{xy} + 2cnp\rho_{zy}}{x^2 + z^2 k^2 + m^2 + 2kxz\rho_{bc} + 2mx\rho_{bd} + 2kmz\rho_{cd}} \quad (2.5)$$

Without further restrictions, this bias can be any real number.⁹ Anticipating the model application below, consider the following empirically-relevant restriction: let the data predictions be nonnegative, the revision errors be uncorrelated, and the parameter shocks be nonnegatively correlated. In this case the bias is positive.

Discussion These simple examples capture the intuition that a researcher's estimate of an agent's model uncertainty is biased by ignoring data uncertainty. The bias is a function of the true variances of the parameter shocks and data errors, the true correlation between data errors, and the perceived correlation between parameter shocks. In the simplest cases, this bias is positive. In

⁹For instance: if all correlations are zero, it is positive; if $\rho_{bc} = \rho_{bd} = \rho_{cd} = 0$, $\rho_{xz} = \rho_{xy} = \rho_{yz} = -1$, and $b = c = n = p = 1$, it is negative.

more complicated cases, the bias may be anything, but under a particular restriction given above the bias is positive. The point of this section is that data uncertainty cannot necessarily be ignored without consequence. To evaluate the scale of the consequence in practice, we turn to the following application.

3 Federal Reserve Model

The question of why inflation rose and fell so dramatically in the United States between 1973 and 1984 has received much attention. One strand of research emphasizes the evolution of the Federal Reserve’s beliefs over this time period. Representing Fed beliefs as parameters in its model, Sargent, Williams, and Zha (2006) reverse engineers an optimal control framework that explains the time path of inflation very well. In light of the previous section, it is worth asking if ignored data uncertainty has an impact on the model’s estimates. We will see in Section 4 that the clear answer to this question is: yes, very much.

3.1 Model

Sargent, Williams, and Zha (2006) assume that the Federal Reserve’s economic model can be usefully approximated by a Philips curve with time-varying parameters. By specifying that the Fed believes the parameters follow a random walk we introduce persistent model uncertainty, as discussed in Cooley and Prescott (1976) and Primiceri (2005). In this context, inflation is the solution to an optimal control problem with a law of motion that changes according to the evolution of filtered parameter estimates. The Federal Reserve’s objective remains the same while new data alter its best estimate of the effects of its actions, represented by the Philips curve trade-off.

The direct effect of Fed activity is the rate of inflation in the economy – it sets inflation up to some exogenous shock beyond its control. This control shock could be thought of as unpredictable market reaction to Fed policy. Therefore the annual inflation rate π_t is

$$\pi_t = x_{t-1} + \frac{1}{\zeta_1} \omega_{1t} \tag{3.1}$$

where x_{t-1} is the part of inflation controllable by the Federal Reserve using information through time $t - 1$, and $\omega_{1t} \sim iid(0, 1)$ is the exogenous control shock.

The Fed uses a Philips curve to understand the relationship between unemployment and inflation. However, the Fed is always uncertain of its estimated model – a way of accomplishing this is

by assuming the parameters follow a random walk:

$$u_t = \alpha'_{t-1} \begin{pmatrix} \pi_t \\ \pi_{t-1} \\ u_{t-1} \\ \pi_{t-2} \\ u_{t-2} \\ 1 \end{pmatrix} + \frac{1}{\zeta_2} \omega_t \equiv \alpha'_{t-1} \Phi_t + \frac{1}{\zeta_2} \omega_{2t} \quad (3.2)$$

$$\alpha_t = \alpha_{t-1} + \Lambda_t \quad (3.3)$$

where $\omega_{2t} \sim iid(0, 1)$ and Λ_t is a vector with $\mathbb{E}(\Lambda_t) = 0$, $\mathbb{E}(\Lambda_t \Lambda_t') = \mathbf{V}$, and $\mathbb{E}(\Lambda_t \omega_{2t}) = \mathbf{0}$.

The reason why the Fed estimates the relationship between unemployment and inflation is because the Fed has inflation and unemployment targets. The objective function, which Sargent (1999) calls the Phelps problem, is written

$$\min_{\{x_{t-1+j}\}_{j=0}^{\infty}} \hat{\mathbb{E}}_t \sum_{j=0}^{\infty} \delta^j \left((\pi_{t+j} - \pi^*)^2 + \lambda (u_{t+j} - u^*)^2 \right) \quad (3.4)$$

where $\delta \in (0, 1)$ is a time discount factor, $\lambda > 0$ gives the Fed's relative weighting of its two objectives, and $\pi^*, u^* \geq 0$ are inflation and unemployment targets. $\hat{\mathbb{E}}$ is expectation with respect to the probability model formed by equations (3.1), (3.2), and (3.3). Because the parameters follow a random walk whose steps are independent of everything else, the Fed's estimate of α_{t-1} is also its estimate of α_{t+j} , $\forall j \geq 0$. Hence, the time t solution to the dynamic programming problem {(3.4) s.t. (3.1), (3.2), (3.3)} is found after plugging the time t estimate of α_{t-1} into the law of motion (3.2) for all $j \geq 0$. I set parameters in line with Sargent, Williams, and Zha (2006): $\delta = .9936$, $\lambda = 1$, $i^* = 2$, $u^* = 1$. They note that the results are unaffected by letting u^* be closer to typical "natural unemployment" rates and I have confirmed that this is indeed the case for both the model with and without data uncertainty.

3.2 Without Data Uncertainty

If the Fed holds no data uncertainty, then the model is completed by assuming that the Fed observes the true values of both inflation and unemployment each period.¹⁰ If this is the case, the Fed estimates the relationships (3.2) and (3.3) through a linear filtering problem whose solution is given by the Kalman filter. The notation $\sigma(\cdot)$ denotes the information set (σ -algebra) formed

¹⁰One might suppose that an Orphanides (2001)-type critique might matter here: this critique would say that the model should use preliminary inflation and unemployment rate data, but would not allow the Fed's filtering problem to recognize the fact that data is revised. In fact, this does not substantially change the problems I discuss below – see Appendix A.IV.2.

by random variables within the parentheses. Let $\mathbb{E}(\boldsymbol{\alpha}_t|\mathcal{I}_s) \equiv \mathbf{a}_{t|s}$ and $\mathbb{V}ar(\boldsymbol{\alpha}_t|\mathcal{I}_s) \equiv \mathbf{P}_{t|s}$ for $\mathcal{I}_s \equiv \sigma(u_1, \pi_1, \dots, u_s, \pi_s)$. Given initial conditions $\mathbf{a}_{1|0}$ and $\mathbf{P}_{1|0}$, the Kalman updating occurs using the formulae:

$$\mathbf{a}_{t+1|t} = \mathbf{a}_{t|t-1} + \frac{\mathbf{P}_{t|t-1} \boldsymbol{\Phi}_t (u_t - \boldsymbol{\Phi}'_t \mathbf{a}_{t|t-1})}{\left(\frac{1}{\zeta_2}\right)^2 + \boldsymbol{\Phi}'_t \mathbf{P}_{t|t-1} \boldsymbol{\Phi}_t} \quad (3.5)$$

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} - \frac{\mathbf{P}_{t|t-1} \boldsymbol{\Phi}_t \boldsymbol{\Phi}'_t \mathbf{P}_{t|t-1}}{\left(\frac{1}{\zeta_2}\right)^2 + \boldsymbol{\Phi}'_t \mathbf{P}_{t|t-1} \boldsymbol{\Phi}_t} + \mathbf{V} \quad (3.6)$$

3.3 Including Data Uncertainty

The modification I make to the model is motivated by the following. Policy-makers base decisions on their model of the economy. Data revisions alter the statistics that inform the policy-makers' model. Hence, the existence of revisions implies that a savvy policy-maker associates some uncertainty to the latest observations of the most recent data vintage. I propose that the policy-maker optimizes accordingly and discipline my analysis with actual evidence on the characteristics of that data uncertainty.¹¹

Graphs of the data show that while the unemployment rate experiences revisions, the CPI inflation rate actually does not. Therefore I only model revisions on unemployment and assume 12-month-ended inflation rates are not subject to revision.¹² Hence, for each month's true unemployment rate u_t we have: a preliminary observation made in month t denoted u_t^0 ; the next observation made in month $t+1$ denoted u_t^1 ; the observation made in month $t+2$ denoted u_t^2 ; and so on. Define the *revision* or *revision error* as

$$\nu_{t+i}^i = u_t^i - u_t^{i-1} \quad , \quad i > 0$$

To reduce the problem's dimension, every revision does not explicitly enter the model and a final vintage horizon $f = 72$ months is chosen.¹³ Therefore, for each u_t I assume that revisions are

¹¹As will be apparent, the key aspect of data uncertainty in this model is that the data is observed with error: one could then disregard data revisions, instead assume data measurement error. I explicitly consider data revisions for two reasons. One, data revisions are evidence that motivate the idea that data is measured with error; two, data revisions have demonstrable statistical properties that discipline data uncertainty's role. This second point is made clearer in Section 4.2.

¹²Some in the Federal Reserve system have noted that policy-makers hold more uncertainty over the inflation rate than the unemployment rate. I am extremely sympathetic to this criticism, but continue to ignore inflation rate uncertainty for the following reason: I only have real-time data on the CPI, which is not revised. Some part of policy-maker's inflation rate uncertainty has to do with the *theoretical issues* concerning inflation, including the issue of Core versus non-Core measures, which are distinct from the data uncertainty I focus on here. Nevertheless, it should be noted that PCE and GDP deflator inflation are subject to revision.

¹³The results presented below are not sensitive to this. In fact, changing f to 36, which accords with the horizon

possible one month, two months, three months, and 72 months later.¹⁴ By grouping all revisions past the third into the final revision, we are relabeling the sum $\sum_{k=4}^f \nu_{t+k}^k$ as simply ν_{t+f}^f . This assumption reduces the dimension of the (augmented) state vector and has no effect on the results.

Therefore the unemployment measurement vector is

$$\begin{bmatrix} u_t^0 \\ u_t^1 \\ u_t^2 \\ u_t^3 \\ u_t^f \\ u_{t-f} \end{bmatrix} = \begin{bmatrix} u_t - \nu_{t+f}^f - \nu_{t+3}^3 - \nu_{t+2}^2 - \nu_{t+1}^1 \\ u_{t-1} - \nu_{t-1+f}^f - \nu_{t-1+3}^3 - \nu_{t-1+2}^2 \\ u_{t-2} - \nu_{t-2+f}^f - \nu_{t-2+3}^3 \\ u_{t-3} - \nu_{t-3+f}^f \\ u_{t-f} \end{bmatrix} \quad (3.7)$$

In any modification that accounts for both data revision and time-varying parameters, both the latent economic quantities (u_t) and the time-varying parameters (α_t) are state variables. Therefore, the state transition equation is nonlinear:

$$\beta_t = \mathbf{g}_t(\beta_{t-1}, \eta_t) \quad (3.8)$$

Ignoring trivial equations that merely shift the position of elements from one state vector to the next, (3.8) is the compact notation for the system of equations

$$\pi_t = x_{t-1}(\cdot) + \frac{1}{\zeta_1} \omega_{1t} \quad (3.9)$$

$$u_t = \alpha'_{t-1} \Phi_t + \frac{1}{\zeta_2} \omega_{2t} \quad (3.10)$$

$$\alpha_t = \alpha_{t-1} + \Lambda_t \quad (3.11)$$

$$\nu_{t+2}^2 = \tilde{\nu}_t^2$$

$$\nu_{t+3}^3 = \tilde{\nu}_t^3$$

$$\nu_{t+f}^f = \tilde{\nu}_t^f$$

where

$$\begin{pmatrix} \tilde{\nu}_t^2 \\ \tilde{\nu}_t^3 \\ \tilde{\nu}_t^f \end{pmatrix} \sim iid(\mathbf{0}, \tilde{\mathbf{V}})$$

of historical values included in the Greenbook, delivers similar results while increasing f does very little. It is only a small f such as 12 or below – where most revisions are ignored and so data is not very uncertain – that affects the results, pushing them back towards the model that ignores data uncertainty altogether. It is unimportant to the results whether $u_t^f = u_t$ or $u_t^f = u_t + error$.

¹⁴The first revision allows for inference on the preliminary report of unemployment. The middle two revisions allow for inference on the latent value of the two lags of unemployment that enter the policy rule emerging from the dynamic programming problem. The final possible revision, which is the largest, acknowledges that data revision happens even after two months.

(3.9), (3.10), and (3.11) repeat (3.1), (3.2), and (3.3), respectively. Note that x has been written as a function in (3.9) in order to point out that it will be a policy rule depending on the best estimate of β_t . Furthermore, (3.9) is an element of the vector Φ_t in (3.10). Hence, (3.9) and (3.10) demonstrate the nonlinear parts of the transition equation.

Turning to the observation equation, the observed vector is

$$\mathbf{y}_t = (\pi_t, u_t^0, u_t^1, u_t^2, u_t^3, u_t^f)'$$

so that the state is measured

$$\mathbf{y}_t = \mathbf{H}\beta_t + \varepsilon_t = \mathbf{H}\beta_t + \begin{pmatrix} 0 \\ \tilde{\nu}_t^1 \\ \mathbf{0} \end{pmatrix} \quad (3.12)$$

where $\tilde{\nu}_t^1 \sim i.i.d. (0, (1/\zeta_\varepsilon)^2)$. \mathbf{H} gives the linear combinations written in (3.7).

3.4 Estimation

While (3.12) is linear, (3.8) is nonlinear. The Extended Kalman filter approximates the state space model using a Taylor-expansion about the linear prediction of the state, as suggested by Anderson and Moore (1979) and following Tanizaki (1996). Considering (3.9)-(3.11), the second-order expansion completely represents the function; however, I have found little difference in practice between the first-order and second-order expansions and use the former since it does not require computing second derivatives numerically. Having a method of approximating the optimal predictions and updates that relies only on matrix multiplication makes the entire estimation procedure computationally reasonable. In the interest of exposition, discussion of the Extended Kalman filter and the likelihood is put in Appendix A.II.

The parameter estimated is

$$\Psi \equiv \left(\zeta_1, \text{vech}(\text{Chol}(\mathbf{V}))', \text{vech}(\text{Chol}(\mathbf{P}_{1|0}))' \right)'$$

where $\text{Chol}(\cdot)$ is the Cholesky factor of positive definite matrix \mathbf{A} such that $\text{Chol}(\mathbf{A})\text{Chol}(\mathbf{A})' = \mathbf{A}$.¹⁵

Because of Ψ 's large dimension, I follow Sargent, Williams, and Zha (2006) and use a Bayesian empirical method discussed in Appendix A.IV. This procedure assumes the shocks are Gaussian.

¹⁵A key point that $\tilde{\mathbf{V}}$ and ζ_ε are *not* estimated is addressed below in Section 4.2.

From the simulated posterior distribution I report medians as my point estimates and quantiles as probability intervals for the parameters. Following Sargent, Williams, and Zha (2006), $\mathbf{a}_{1|0}$ is set to a regression estimate from presample data and ζ_2 is set to be 59, indicating the Fed believes its Philips curve, if the parameters were known, would deliver forecasts with one-tenth the RMSE of naive random walk forecasts; see Appendix A.I for a detailed discussion.¹⁶

Given a prior $p(\Psi)$ and the likelihood $\mathcal{L}(\mathcal{Y}_T|\Psi)$ (in Appendix A.II), the posterior distribution is

$$p(\Psi|\mathcal{Y}_T) \propto \mathcal{L}(\mathcal{Y}_T|\Psi)p(\Psi) \tag{3.13}$$

I sample from (3.13) using a Metropolis algorithm with random walk proposals (cf. Robert and Casella (2004)).¹⁷

4 Empirical Results

The outline of this section is as follows. The first subsection presents the estimation results for the model without data uncertainty, which are essentially identical to those in Sargent, Williams, and Zha (2006). I describe three problems that emerge from these results. The next subsection explains the importance of disciplining my model with the observed characteristics of data revisions. The significance of this discipline is straightforward after one understands the problems that must be addressed. The last subsection presents the estimation results for the model including data uncertainty which fixes the fore mentioned problems.

4.1 Without Data Uncertainty

Data on inflation and the civilian unemployment rate for ages 16 and older comes from the BEA and BLS, respectively, as reported in December 2003.¹⁸ The data begins in 1960 and I end the

¹⁶The parameter ζ_2 is unidentified in in the model without data uncertainty and so Sargent, Williams, and Zha (2006) normalize ζ_2 such that $\frac{1}{\zeta_2}$ is one-tenth the standard deviation of the shock in an additional equation for the “true DGP” of the unemployment rate. As Carboni and Ellison (2008) discuss, apart from this choice the “true DGP” equation does not influence the belief-generating mechanism I have described; thus, it is not part of my model. In the interest of making a clear comparison to its results, I take the value of ζ_2 from Sargent, Williams, and Zha (2006).

¹⁷I must use a accept/reject simulation technique because, due to the effects of Ψ on the whole sequence of forecasts, the form of (3.13) is not known. Further details are in Appendix A.IV.

¹⁸The inflation rate data, following Sargent, Williams, and Zha (2006) is the annual rate of change of the seasonally-adjusted Personal Consumption Expenditure chain price index from the BEA, as reported in December 2003. I use PCE inflation here in the interest of making a direct comparison although I must use real-time CPI inflation for the model including data uncertainty. Estimates of the model without data uncertainty using current vintage non-seasonally-adjusted or seasonally-adjusted CPI inflation are very similar and so this does not drive the difference

Table 1: PARAMETER ESTIMATES, WITHOUT DATA UNCERTAINTY

<hr/> <hr/>						
ζ_1 : 4.24 (3.96, 4.47)						
<hr/> <hr/>						
\mathbf{V} : standard deviations and <i>correlations</i>						
0.2871	-0.9517	0.1860	0.9708	-0.1256	-0.2835	
	0.2844	0.002	-0.9976	0.2953	0.47	
		0.1723	0.0426	0.9442	0.8187	
			0.1769	-0.256	-0.4346	
				0.2274	0.8712	
					5.0817	
<hr/> <hr/>						

Notes: at median of posterior distribution. For top panel, ζ_1^2 is the precision of ω_{1t} , the additive shock to the Fed’s inflation control, and 95% probability intervals in parentheses. The bottom array is comprised as follows: the main diagonal are the square roots of the main diagonal of \mathbf{V} ; the off-diagonal elements are the correlations derived from \mathbf{V} ; \mathbf{V} is the covariance matrix of the Λ_t shock to the time-varying parameters α_t . The vector $\Phi_t = (\pi_t, \pi_{t-1}, u_{t-1}, \pi_{t-2}, u_{t-2}, 1)'$ multiplies α_t .

sample at December 1995.¹⁹ Table 1 shows estimates from 75,000 draws derived from 100,000 MCMC iterations where the first 25% are burned in hopes that the Markov Chain has, for practical purposes, converged to its ergodic distribution.²⁰ The estimates in Table 1 are virtually identical to those of Sargent, Williams, and Zha (2006).²¹

Figure 1 shows the predicted inflation control choices. Figure 1 shows the Fed *choosing* to set inflation high in the two high-inflation episodes of the mid 1970s and early 1980s. With ζ_1 estimated at about 4.24, the standard deviation of the ω_{1t} is around 0.2, reflecting the Fed’s belief that it has rather tight control of inflation.

The Philips curve beliefs $a_{t-1|t-1}$ are used to forecast the next month’s unemployment rate for any inflation control setting. According to the model, the Fed sets inflation with this forecast in mind. Therefore an important aspect of the model-predicted Fed beliefs are what they deliver in terms of unemployment forecasts: these are plotted in Figure 2.²² These forecasts have some neg-

between the models’ estimates; see Appendix A.IV.1.

¹⁹The results are the same if the data runs through December 2003; I end the sample at 1996 in order to ensure comparability between all the models I have estimated for robustness (one of which assumes final data is seen 10 years after the fact) and because I consider the model to be descriptive of Fed beliefs only through the 1980s.

²⁰Sargent, Williams, and Zha (2006)’s results come from a sequence of 50,000 draws with an unspecified burn-in interval.

²¹They also accord with results in Carboni and Ellison (2008) who also estimate this model. Appendix A.IV contains the estimates of $\mathbf{P}_{1|0}$ – this parameter is far less important than \mathbf{V} since the influence of $\mathbf{P}_{1|0}$ on the gain (which imputes shocks from forecast errors) dies out within a dozen or so periods, while \mathbf{V} dictates the steady-state gain.

²²The information displayed in Figure 2 is implied by Sargent, Williams, and Zha (2006)’s Figure 7. The difference is that in their Figure 7 the inflation control is replaced by the “Ramsey” inflation rate policy of 2, the graph is shifted downward by the estimate of the true natural rate of unemployment, and the parameter estimate used for

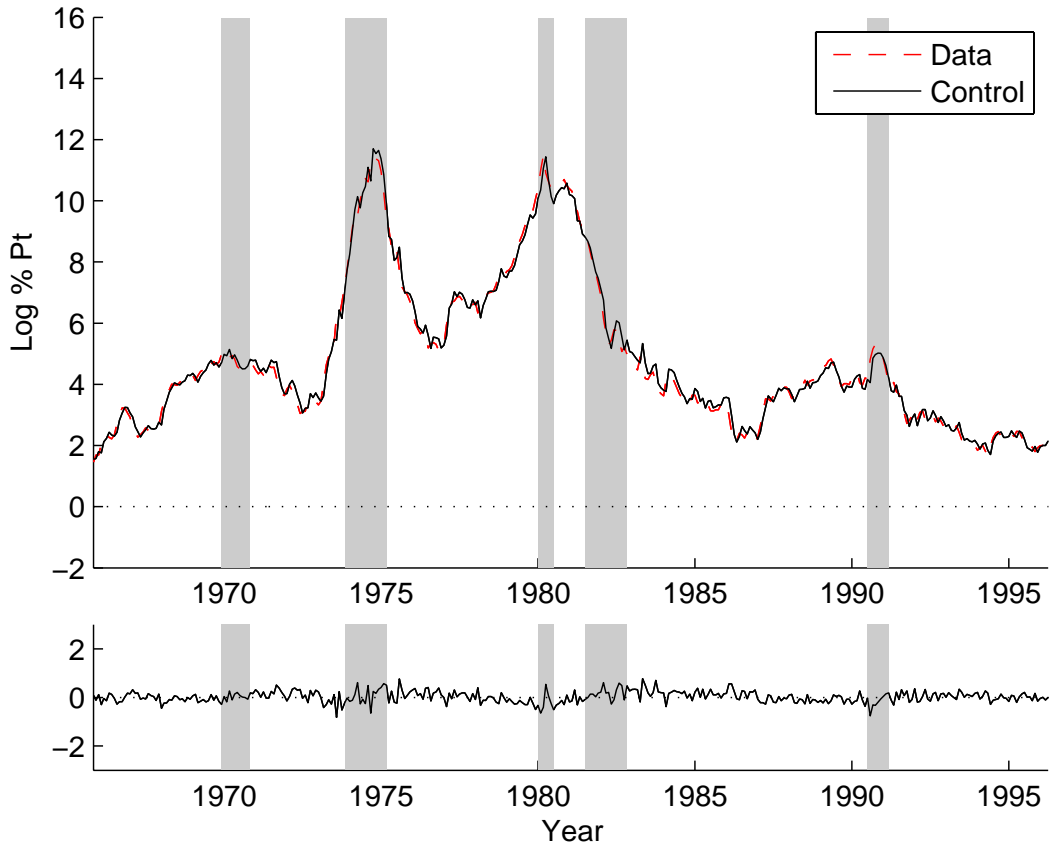


Figure 1: ACTUAL INFLATION VS. FED CONTROL, WITHOUT DATA UNCERTAINTY

Notes: *Top panel:* Actual inflation and model predicted Fed inflation control. *Bottom panel:* prediction errors. NBER recessions shaded. Figures 1 and 3 are on the same scale.

ative bias and a Root Mean Squared Error of 3.3 percentage points. Roughly speaking, every four months the Fed expects its month-ahead unemployment rate forecasts to be off by 3.3 *percentage* points.

An important aspect of these unemployment rate forecasts is how well they explain actual Federal Reserve unemployment rate forecasts. There is evidence of the latter from the Greenbook forecasts over this time span.²³ An appropriate statistical test of the similarity between the model-predicted forecasts and the Greenbook forecasts is that of Diebold and Mariano (1995). Their statistic S_1 is a two-sided test of the null hypothesis that the model's unemployment forecasts have

time t is the updated estimate $\mathbf{a}_{t|t}$ (which is identically $\mathbf{a}_{t+1|t}$).

²³Appendix A.III discusses these forecasts and provides more details on the test statistic.

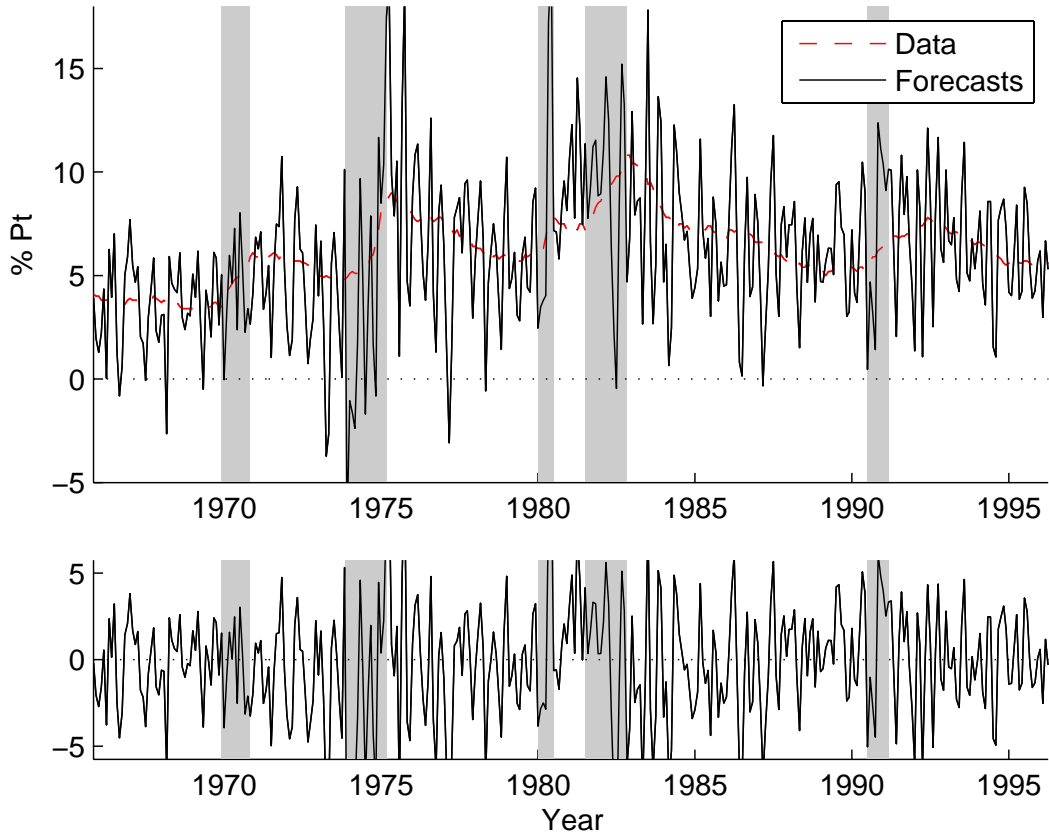


Figure 2: ACTUAL UNEMPLOYMENT VS. FED FORECASTS, WITHOUT DATA UNCERTAINTY

Notes: *Top panel:* step-ahead unemployment forecasts come from the Philips curve (3.2) using the Fed’s inflation setting and actual unemployment and inflation data. *Bottom panel:* predicted forecast errors. NBER recessions shaded. Figures 2 and 4 are on the same scale.

accuracy equal to the Greenbook forecasts and S_1 has an asymptotic standard normal distribution. Here $|\hat{S}_1| = 2.5044$, so we can reject the null of equal forecasting accuracy at the 99% level. This means that in terms of accuracy the model-predicted forecasts are statistically different than actual Greenbook forecasts.

Given the large value of ζ_2 (equivalently, the small variance of ω_{2t}), volatile and inaccurate unemployment forecasts are evidence of greatly varying parameters. Hence the estimate of the parameter shock’s covariance matrix is large. In particular, notice that the estimated $V^{(6,6)}$ implies that the Fed believes that every month the Philips curve’s intercept experiences an i.i.d. shock with a standard deviation of about 5 unemployment rate points. This approximately means the

Fed believes that every month there is a 30% chance the natural rate of unemployment jumps by 5 percentage points in either direction, even if inflation is kept at target. In other words, the Fed believes its model of the world is wildly unstable.

These results render the optimal policy story very problematic. On the one hand, the estimates imply the Fed regarded its unemployment forecasting tool as subject to large unpredictable shifts that cause its forecasts to be erratic and inaccurate. On the other hand, the model's main point is that the Fed's estimated Philips trade-off was motivation for inflationary policy in hopes of lowering unemployment. Given the large estimate of V and its poor forecasting performance, it is implausible that the Fed believed its Philips curve enough to take on the pain of high inflation in hopes of decreasing unemployment. Moreover, even if the Fed did believe its estimated Philips curve enough to use it as basis for policy, we will see below that the evolution of the Philips curve trade-off does not explain inflation's rise *and* fall.

4.2 Observing Revision Errors

At this point it is clear that my modification introduces a new shock called the data revision error. Moreover, note that data revision errors and time-varying parameter shocks both affect the Fed's unemployment forecast performance: the Fed would expect these forecasts to be poor if the data is imprecisely observed or the time-varying parameters are jumping around a lot. This idea emerges in the Fed's filtering problem that decomposes unemployment forecast errors into predicted parameter shocks and predicted data revision errors.²⁴ Thus, by simply introducing another shock into the model, we might soak up some of the force driving the previous section's results, where large forecast errors went hand-in-hand with large parameter shocks. However, we would be doing so by simply adding another latent variable, and the degree to which the forecast errors were attributed to parameter shocks versus data revision errors would be mostly due to our priors on their respective sizes.

I avoid this problem by specifying the properties of the data revision errors instead of estimating them. The variances and covariances of the revision errors $(\tilde{V}, \zeta_\varepsilon)$ are set to values independently estimated on the data; basically, there is no correlation between revision errors and the largest

²⁴I am ignoring the role of the additive shock ω_{2t} because its variance is set so small and, in any case, is identical across the models.

revision error by far is the final one whose standard deviation is about 0.11.²⁵ This is tantamount to imposing that the Fed knows the data uncertainty surrounding the unemployment rate in the same way it knows its own objective function parameters. This assumption is reasonable since data revisions are observable and their statistical properties are easily known. This way, model estimation does not deliver results that *necessarily* drive down the size of parameter shocks by attributing an exaggerated role to data uncertainty. Instead, the model takes as given and observable the characteristics and realizations of the data revision errors, and estimates the size and time path of the parameter shocks in light of this data.

4.3 Including Data Uncertainty

Multiple vintages of real-time data on inflation and the civilian unemployment rate for ages 16 and older comes from the ALFRED archive maintained by the Federal Reserve Bank of St. Louis, downloaded in 2007.²⁶ Table 2 reports estimates from 700,000 MCMC iterations from two separate

Table 2: PARAMETER ESTIMATES, INCLUDING MODEL UNCERTAINTY

	ζ_1 :	2.23	(2.21, 2.25)			
		<i>V</i> : standard deviations and <i>correlations</i>				
0.0192	0.1749	-0.523	0.5294	0.3461	-0.8742	
	0.0227	-0.2719	-0.69	0.329	-0.1155	
		0.0478	-0.1357	-0.9727	0.2646	
			0.0242	-0.0012	-0.3808	
				0.0436	-0.0682	
					0.1953	

Notes: at median of posterior distribution. For top panel, ζ_1^2 is the precision of ω_{1t} , the additive shock to the Fed’s inflation control, and 95% probability intervals in parentheses. The bottom array is comprised as follows: the main diagonal are the square roots of the main diagonal of \mathbf{V} ; the off-diagonal elements are the correlations derived from \mathbf{V} ; \mathbf{V} is the covariance matrix of the $\mathbf{\Lambda}_t$ shock to the time-varying parameters $\mathbf{\alpha}_t$. The vector $\mathbf{\Phi}_t = (\pi_t, \pi_{t-1}, u_{t-1}, \pi_{t-2}, u_{t-2}, 1)'$ multiplies $\mathbf{\alpha}_t$.

runs of 400,000 with different initial conditions where the first 50,000 of each run is burned.

Figure 3 shows that the Fed’s inflation control explains the rise and fall of American inflation. The estimated standard deviation of the inflation control shock is 0.4 and the fit of the model-predicted inflation to actual inflation is a little poorer than in Section 4.1 and Sargent, Williams,

²⁵The matrix $\tilde{\mathbf{V}}$ is diagonal with main diagonal (0.0007, 0.0004, 0.0121)’ and ζ_ε is 45.273.

²⁶I use non-seasonally-adjusted CPI inflation because real-time PCE inflation is not available for the time span under consideration. If we assume current vintage PCE inflation data was known in real-time and underwent no revisions, the results presented here are virtually identical.

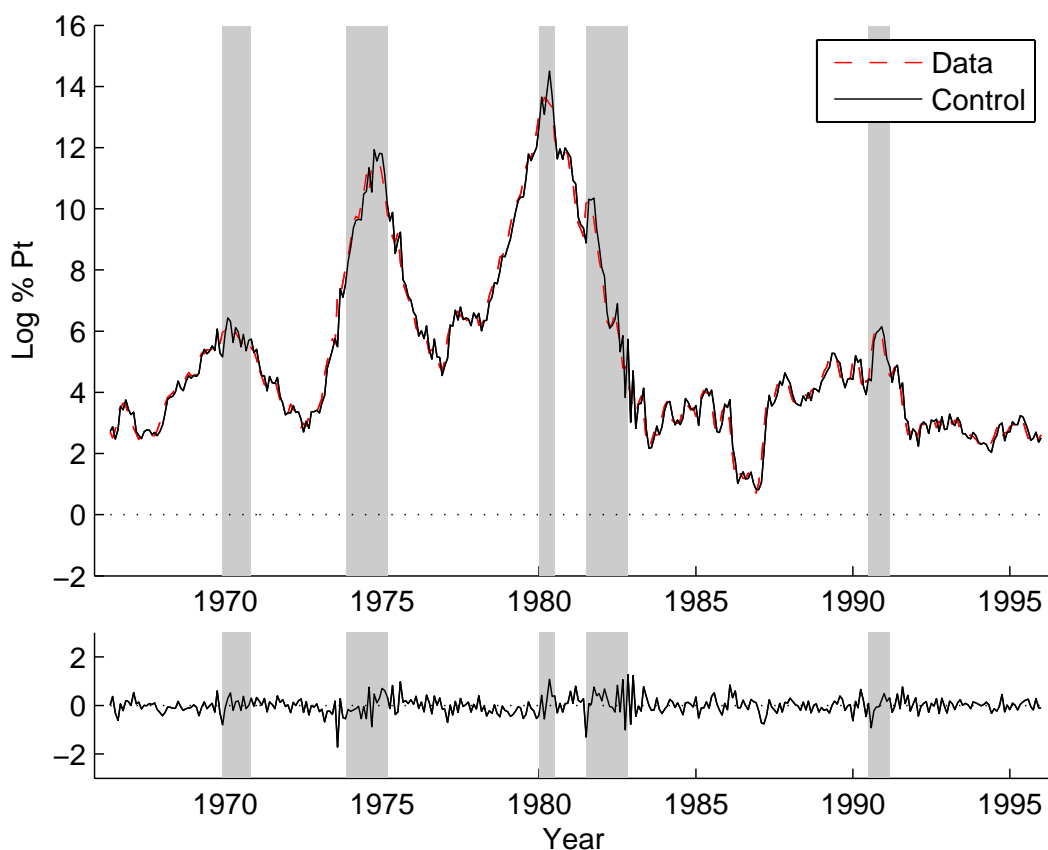


Figure 3: ACTUAL INFLATION VS. FED CONTROL, INCLUDING DATA UNCERTAINTY

Notes: *Top panel:* Actual inflation and model predicted Fed inflation control. *Bottom panel:* prediction errors. NBER recessions shaded. Figures 1 and 3 are on the same scale.

and Zha (2006). However, since the point of the model is to explain low frequency movements in inflation (the Great Inflation), I argue that this deterioration in high-frequency fit is not a problem.

Turning now to the Fed’s unemployment rate forecasts in Figure 4, we find a far different picture than in the model without data uncertainty. The Fed’s forecasts are considerably more accurate than before, with no bias and a RMSE of 0.24 percentage points.²⁷

Again, seeing as the model forecasts are intended to predict the Fed’s actual forecasts, we can directly compare them to Greenbook unemployment rate forecasts. Using the Diebold and Mariano (1995) test of equal accuracy between model forecasts and Greenbook forecasts, $|\hat{S}_1| = 0.4858$ and we accept the hypothesis that the model forecasts and Greenbook forecasts are equally-accurate.

²⁷The other measurement equation forecasts are pictured in Appendix A.IV.

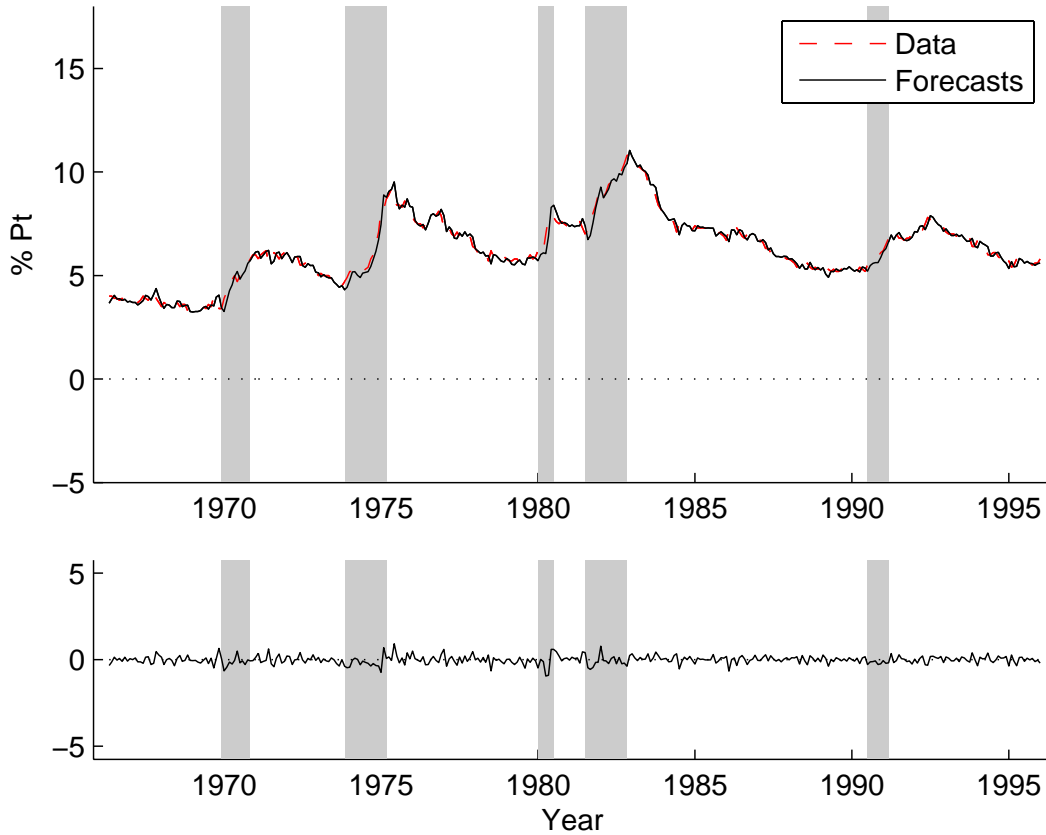


Figure 4: ACTUAL UNEMPLOYMENT VS. FED FORECAST, INCLUDING DATA UNCERTAINTY

Notes: *Top panel:* step-ahead unemployment forecasts come from the Philips curve (3.2) using the Fed’s inflation setting and actual unemployment and inflation data. *Bottom panel:* predicted forecast errors. NBER recessions shaded. Figures 2 and 4 are on the same scale.

Therefore in terms of accuracy the model-predicted unemployment rate forecasts are statistically indistinguishable from Fed unemployment rate forecasts.

The estimate of \mathbf{V} in Table 2 is much smaller than before. For instance, the estimated $V^{(6,6)}$ implies that the Philips curve’s intercept has a monthly shock with a standard deviation of about $\frac{1}{5}$ as opposed to the 5 points estimated with data uncertainty left out.²⁸ Roughly speaking, the shocks are smaller by about 8 times for inflation parameters, 5 times for unemployment parameters,

²⁸It might be suggested that one could adopt part of the model without data uncertainty *ad hoc* and include unemployment forecast errors as part of a modified version of an estimation procedure resembling the model with data uncertainty. Doing so might by itself lead to smoother TVP estimates and better fitting inflation and unemployment forecasts. Looking into this, I found that while the unemployment forecasts errors do improve, the fit of the model-predicted inflation control deteriorates greatly; see Appendix A.IV.3. This is somewhat similar to, but without the success of, what Carboni and Ellison (2008) does.

and 20 times for the constant parameter. The Fed is now confident that its Philips curve is not hopelessly unstable. It is now plausible that the Fed believes (3.2) and sets inflation in light of an existing Philips curve trade-off.

Let us turn now to the predicted evolution of the Fed's beliefs about the Philips curve trade-off. As seen in the top panel of Figure 5, the trade-off in Sargent, Williams, and Zha (2006) experiences a large jump between 1973 and 1975 which explains the great rise in inflation over those years. Thereafter the trade-off shows a gradual decline, with no sharp activity around the disinflation of the early 1980s.²⁹ But this does not bear out the main story, which is that the evolution of the Philips curve trade-off led to the rise *and fall* of inflation. In the model without data uncertainty, the dramatic fall of inflation from 14.4% to 2.2% between 1980 and 1984 occurs without any concurrent sharp change in the Fed's beliefs.

On the other hand, consider the bottom panel predicted by the model including data uncertainty. I predict a drastic drop in the Philips curve trade-off starting around 1980. As this perceived trade-off falls by 98% off its peak, inflation falls by than 80% off its peak, and both bottom out in 1984. Moreover, the lull in inflation in the late 1970s corresponds to the fall in the negative Philips curve relationship starting in 1975.³⁰ Thus, the model including data uncertainty describes a strong and consistent connection between the Fed's inflation control and the Fed's beliefs about the Philips curve trade-off.

In sum, we have explicitly allowed for some data uncertainty that the Fed might have held about the unemployment situation. Since this data is revised, it seems natural to suppose the Fed wouldn't regard real-time observations as exact. By so doing, we drastically but constructively change the results in Sargent, Williams, and Zha (2006). Now the model reverse engineers a plausible story that the Fed's evolving Philips curve beliefs were directly tied to its control of inflation.

5 Why Data Uncertainty Matters Here

An obvious question is, how can such small data revision errors cause such large changes in the model performance and estimates? The answer lies in understanding that the errors are not small

²⁹The estimated trade-off evolution in Carboni and Ellison (2007) and Carboni and Ellison (2008) is qualitatively similar, the latter with a much smaller magnitude but a very similar shape.

³⁰Over the sample period there is a strong correlation of -0.75 between the low-frequency fluctuations of the trade-off and inflation. Low-frequency fluctuations as those with a period between 1 and 10 years, as estimated using the bandpass filter. For the model without inflation this correlation is -0.14 .

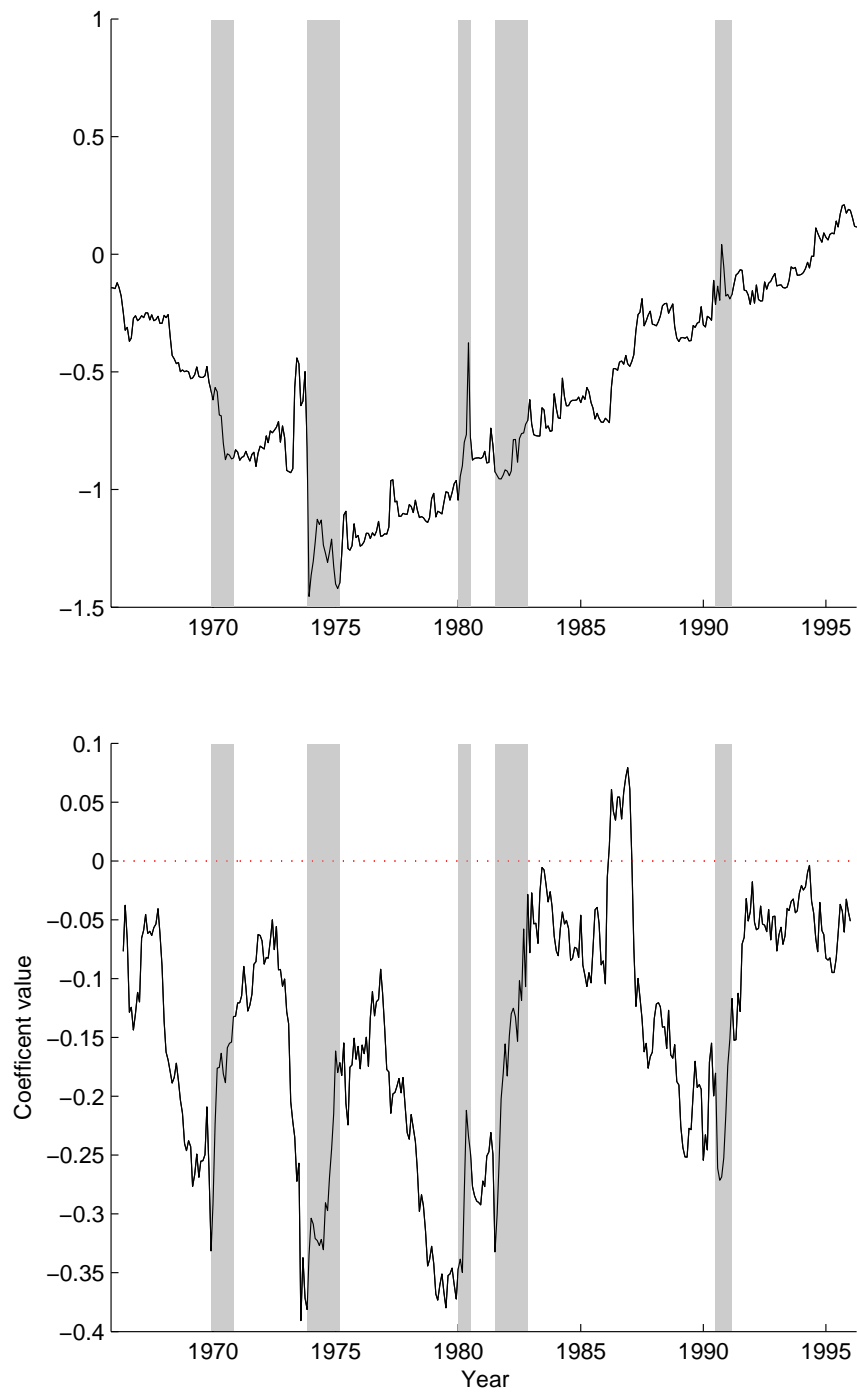


Figure 5: EVOLUTION OF THE PHILIPS CURVE TRADE-OFF

Notes: **Top panel:** time series of Philips curve inflation response estimates from model without data uncertainty. **Bottom panel:** time series of Philips curve inflation response estimates from model including data uncertainty. NBER recessions are shaded.

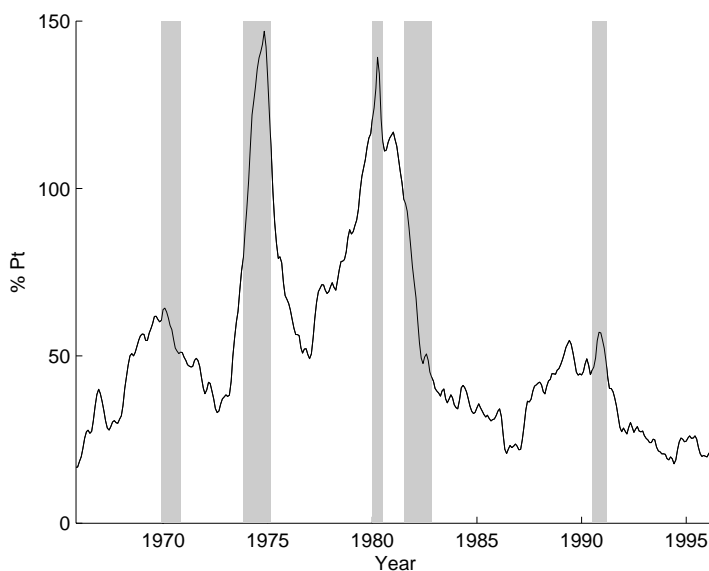


Figure 6: PERCENTAGE BIAS OF ESTIMATED PARAMETER SHOCK SIZE

Notes: at each point in time, bias determined by inflation and unemployment predicted values, perceived correlation and relative size of parameter shocks from model without data uncertainty, and actual relative size of unemployment errors to unemployment parameter shocks from model including data uncertainty. NBER recessions shaded.

in comparison to the unemployment parameter shocks.

In standard deviation terms, the data errors are about three times larger than the shocks to the unemployment parameters. On the other hand, by ignoring data uncertainty the researcher perceives the correlations between unemployment parameter shocks and the other parameter shocks as predominantly positive – see Table 1. This situation is reminiscent of the last example in Section 2 under the particular restriction that data predictions are nonnegative, data errors are uncorrelated, and parameter shocks are nonnegatively correlated. Hence, we might expect that the researcher’s estimated parameter shock size unit would be biased upwards, and indeed this is the case.

Figure 6 shows the bias on estimates of the parameter shock size unit. This bias, which recall is important for decomposing the forecast error into the part contributed by parameter shocks, is time-varying because it depends on the value of each period’s data and time-varying parameter predictions. During the sample period, this bias is always greater than 10% and attains values as high as 146%.

In essence, the overestimate of the unemployment parameters’ shocks leads the model to overestimate the shocks to the inflation and constant parameters in order to explain inflation, and these

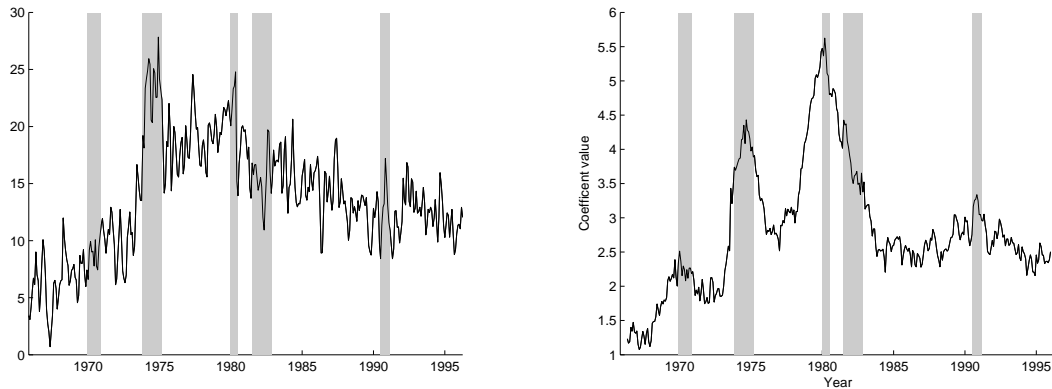


Figure 7: EVOLUTION OF THE PHILIPS CURVE CONSTANT

Notes: *Left panel:* time series of Philips curve constant estimates from model without data uncertainty. *Right panel:* time series of Philips curve constant estimates from model including data uncertainty. NBER recessions are shaded.

greatly varying parameters rot the unemployment rate forecast performance; the explanation of this is as follows. By ignoring the Fed’s data uncertainty, we overestimate the size of the shocks to the parameters on unemployment in the Philips curve (3.2). Because of this, the unemployment beliefs are adjusted too much in response to an unemployment rate forecast error. Recall that these parameters are key ingredients to the optimal policy rule governing inflation. But since inflation over this period is rather persistent, *ceteris paribus* next period’s optimal policy rule should be (to fit the data) somewhat near to this period’s policy rule. Therefore, the size of shocks to the parameters on inflation and the constant are overestimated: this allows them to shift a lot too, period to period, in order to smooth out the path of the policy rule. This is clearly seen in the time path of Philips curve constant, plotted in Figure 7, which moves around a lot in order to fit the policy rule to actual inflation. The constant term belief moves much more sluggishly when data uncertainty is explicitly modeled.

This example suggests that ignored data uncertainty can seriously influence a model’s results. In particular, estimates of the agent’s model uncertainty are positively biased. For this bias, the size of the data uncertainty relative to the relevant model uncertainty – here the size of shocks to the parameters on unemployment – is important. Moreover, the process of fitting the model to the data can exacerbate the issue.

6 Conclusion

This paper analyzes the effects of data uncertainty in models with agents who update their beliefs. Ignoring data uncertainty can seriously bias estimates of agents' model uncertainty. After illustrating this point through simple examples, I apply it to the more sophisticated framework of Sargent, Williams, and Zha (2006). I show that accounting for data uncertainty remedies problems having to do with overestimates of the Fed's model uncertainty and the dissimilarity of model-predicted unemployment rate expectations to actual Greenbook evidence. The Federal Reserve is more sluggish to change its beliefs in response to new data when this data may be revised. Once this is the case, the model predicts that the inflation of the 1970s and 1980s can be strongly tied to evolving beliefs about the Philips curve trade-off between inflation and unemployment.

References

- ANDERSON, B. D. O., AND J. B. MOORE (1979): *Optimal Filtering*. Prentice-Hall.
- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–58.
- ARUOBA, S. B. (2004): “Data Uncertainty in General Equilibrium,” Computing in Economics and Finance 2004 131, Society for Computational Economics.
- BOIVIN, J. (2006): “Has US Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data,” *Journal of Money, Credit, and Banking*, 38(5), 1149–1173.
- CARBONI, G., AND M. ELLISON (2007): “Learning and the Great Inflation,” Discussion paper, University of Warwick.
- (2008): “The Great Inflation and the Greenbook,” Discussion paper, University of Oxford.
- CLARIDA, R., J. GALI, AND M. GERTLER (2000): “Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory,” *The Quarterly Journal of Economics*, 115(1), 147–180.
- COOLEY, T. F., AND E. C. PRESCOTT (1976): “Estimation in the Presence of Stochastic Parameter Variation,” *Econometrica*, 44(1), 167–84.
- CROUSHORE, D., AND T. STARK (2001): “A real-time data set for macroeconomists,” *Journal of Econometrics*, 105(1), 111–130.
- CUNNINGHAM, A., C. JEFFERY, G. KAPETANIOS, AND V. LABHARD (2007): “A State Space Approach To The Policymaker’s Data Uncertainty Problem,” Money Macro and Finance (MMF) Research Group Conference 2006 168, Money Macro and Finance Research Group.
- DIEBOLD, F. X., AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–63.
- ELLIOTT, G., I. KOMUNJER, AND A. TIMMERMANN (2005): “Estimation and Testing of Forecast Rationality under Flexible Loss,” *Review of Economic Studies*, 72(4), 1107–1125.
- EVANS, G. W., AND S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.
- FRIEDMAN, M. (1968): “The Role of Monetary Policy,” *The American Economic Review*, 58(1), 1–17.
- GHYSELS, E., N. R. SWANSON, AND M. CALLAN (2002): “Monetary Policy Rules with Model and Data Uncertainty,” *Southern Economic Journal*, (69), 239–265.
- HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- HOWREY, E. P. (1978): “The Use of Preliminary Data in Econometric Forecasting,” *The Review of Economics and Statistics*, 60(2), 193–200.
- NASON, J. M., AND G. W. SMITH (forthcoming): “The New Keynesian Phillips Curve: Lessons from Single-Equation Econometric Estimation,” *Economic Quarterly*.

- OH, S., AND M. WALDMAN (1990): “The Macroeconomic Effects of False Announcements,” *The Quarterly Journal of Economics*, 105(4), 1017–34.
- ORPHANIDES, A. (2001): “Monetary Policy Rules Based on Real-Time Data,” *American Economic Review*, 91(4), 964–985.
- ORPHANIDES, A., AND J. C. WILLIAMS (2006): “Monetary Policy with Imperfect Knowledge,” *Journal of the European Economic Association*, 4(2-3), 366–375.
- OWYANG, M. T., AND G. RAMEY (2004): “Regime switching and monetary policy measurement,” *Journal of Monetary Economics*, 51(8), 1577–1597.
- PRIMICERI, G. E. (2005): “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies*, 72(3), 821–852.
- RAFTERY, A., AND S. LEWIS (1992): “How Many Iterations in the Gibbs Sampler,” .
- ROBERT, C. P., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*. Springer, 2 edn.
- ROMER, C. D., AND D. H. ROMER (1990): “Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz,” NBER Working Papers 2966, National Bureau of Economic Research, Inc.
- RUNKLE, D. E. (1998): “Revisionist history: how data revisions distort economic policy research,” *Quarterly Review*, (Fall), 3–12.
- SARGENT, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press.
- SARGENT, T. J., N. WILLIAMS, AND T. ZHA (2006): “Shocks and Government Beliefs: The Rise and Fall of American Inflation,” *American Economic Review*, 96(4), 1193–1224.
- SIMS, C. (2007): “Evidence for passive monetary policy pre-1979, comments,” Discussion paper, Federal Reserve Bank of Cleveland conference.
- SIMS, C. A. (2006): “Improving Monetary Models,” Discussion paper.
- TANIZAKI, H. (1996): *Nonlinear Filters: Estimation and Applications*. Springer-Verlag, 2 edn.
- WATSON, M. W., AND R. F. ENGLE (1983): “Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models,” *Journal of Econometrics*, 23, 385–400.
- ZELLNER, A. (1958): “A Statistical Analysis of Provisional Estimates of Gross National Product and its Components, of Selected National Income Components, and of Personal Saving,” *Journal of the American Statistical Association*, 53(281), 54–65.

A Appendix

A.I Federal Reserve Model Comparison

I have dropped Sargent, Williams, and Zha (2006)’s “true” Lucas natural-rate Philips curve from my model presentation. The estimation of the Sargent, Williams, and Zha (2006) Lucas natural-rate Philips curve does not affect the estimates of the belief-formation parameters apart from ζ_2 because the relevant priors are independent and, otherwise, the likelihood does not tie the two together: this relates to Carboni and Ellison (2008)’s statements that including the “true” curve in the overall model gives an ‘informational gain [that] is low’ and that the belief-formation results ‘are robust to embedding’ the model of Section 3.1 ‘in a “true” model.’

Note that ζ_2 is unidentified by the model without data uncertainty. Sargent, Williams, and Zha (2006) deals with this problem by normalizing it such that $\frac{1}{\zeta_2}$ is one-tenth the standard deviation of the shock in the Lucas natural-rate equation. In practice, that paper’s assumption implies that $\zeta_2 = 59.7108$ which means that the Fed thinks that, if it knew the parameters, the Philips curve would forecast unemployment up to an exogenous error with standard deviation $\frac{1}{59.7108} = 0.0167$. That paper describes the assumption as implying that the Fed believes “that the standard deviation of the [its] regression error is smaller by a factor of ten than the standard deviation exogenous unemployment shocks.”

Since the Philips curve is used by the Fed to forecast its affects on unemployment, consider a very common forecasting rule: the random-walk unemployment forecast. I can make an assumption relative to random-walk unemployment forecast errors that is similar to Sargent, Williams, and Zha (2006)’s assumption relative to Lucas natural-rate Philips curve shocks. This assumption is: “the Fed believes that the standard deviation of its Philips curve forecast error is smaller by a factor of ten than the standard deviation of random-walk unemployment forecast errors.” This assumption in practice implies that $\zeta_2 = 54.9753$ since the standard deviation of random-walk unemployment forecast errors is 0.1819. Hence I arrive at choice for ζ_2 that is similar to Sargent, Williams, and Zha (2006)’s, but without the Lucas natural-rate Philips curve.

To ensure that my results are robust to these factors, I have done the following. I have reproduced the results of Sargent, Williams, and Zha (2006)’s entire model. I have estimated the Section 3.2 model while setting $\zeta_2 = 59.7108$ (calculated from Sargent, Williams, and Zha (2006)’s estimate of 35.6538 as the precision of the shock to the Lucas natural-rate Philips curve). And I have estimated the Section 3.2 model by using $\zeta_2 = 54.9753$ based on assumed Fed beliefs as to its forecasting ability relative to random-walk forecasts. The results (parameter estimates, beliefs, inflation choices, and unemployment forecasts) are virtually the same.³¹ In order to provide a clear comparison to previous literature, I simply set $\zeta_2 = 59$.

As an aside, ζ_2 is identified in the model including data uncertainty: the affect of \mathbf{V} on the first derivative of the optimal control policy rule allows that parameter to be pinned down apart from ζ_2 , eliminating the identification problem that exists in the totally linear model without data uncertainty. However, I do not estimate this parameter so as to make as clear as possible my results’ comparison to Sargent, Williams, and Zha (2006).

A.II Extended Kalman Filter and Likelihood

Extended Kalman Filter To implement, we first approximate the state space model by first-order expansions. Let $\mathbb{E}(\beta_t|\mathcal{Y}_s) \equiv \mathbf{b}_{t|s}$ and $\text{Var}(\beta_t|\mathcal{Y}_s) \equiv \Sigma_{t|s}$ for $\mathcal{Y}_s \equiv \sigma(\mathbf{y}_s, \mathbf{y}_{s-1}, \dots)$. The

³¹The value of log-likelihood (multiplied by the prior) for my reproduction of Sargent, Williams, and Zha (2006) is 555.0586, comparable to the 564.92 they find.

expansion of (3.12) about $(\boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t) = (\mathbf{b}_{t|t-1}, 0)$ is exact:

$$\mathbf{h}_t(\boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t) = \mathbf{H}\mathbf{b}_{t|t-1} + \mathbf{H}(\boldsymbol{\beta}_t - \mathbf{b}_{t|t-1}) + \boldsymbol{\varepsilon}_t \quad (\text{A.1})$$

The expansion of (3.8) about $(\boldsymbol{\beta}_{t-1}, \boldsymbol{\eta}_t) = (\mathbf{b}_{t-1|t-1}, 0)$ is approximate:

$$\mathbf{g}_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\eta}_t) \approx \mathbf{g}_{t|t-1} + \mathbf{T}_{t|t-1}(\boldsymbol{\beta}_t - \mathbf{b}_{t-1|t-1}) + \mathbf{R}_{t|t-1}\boldsymbol{\eta}_t \quad (\text{A.2})$$

where

$$\begin{aligned} \mathbf{g}_{t|t-1} &= \mathbf{g}_t(\mathbf{b}_{t-1|t-1}, 0) \\ \mathbf{T}_{t|t-1} &= \left. \frac{\partial \mathbf{g}_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\eta}_t)}{\partial \boldsymbol{\beta}'_{t-1}} \right|_{(\mathbf{b}_{t-1|t-1}, 0)} \\ \mathbf{R}_{t|t-1} &= \left. \frac{\partial \mathbf{g}_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\eta}_t)}{\partial \boldsymbol{\eta}'_t} \right|_{(\mathbf{b}_{t-1|t-1}, 0)} \end{aligned}$$

I motivate the derivation of optimal prediction and updating for the approximating system (A.1) and (A.2) by assuming Gaussian shocks, as in Howrey (1978), Watson and Engle (1983), and Harvey (1989).³² In this case, the relevant conditional expectations have the known forms given below. In particular, I assume

$$\boldsymbol{\eta}_t \sim iid \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \boldsymbol{\varepsilon}_t \sim iid \mathcal{N}(\mathbf{0}, \mathbf{N}), \quad \boldsymbol{\eta}_t \perp \boldsymbol{\varepsilon}_\tau, \forall t, \tau$$

The Extended Kalman Filtering equations are

$$\mathbf{b}_{t|t-1} = \mathbf{g}_{t|t-1} \quad (\text{A.3})$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{T}_{t|t-1}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{T}'_{t|t-1} + \mathbf{R}_{t|t-1}\mathbf{Q}\mathbf{R}'_{t|t-1} \quad (\text{A.4})$$

$$\mathbf{y}_{t|t-1} = \mathbf{H}\mathbf{b}_{t|t-1} \quad (\text{A.5})$$

$$\mathbf{F}_{t|t-1} = \mathbf{H}\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}' + \mathbf{N} \quad (\text{A.6})$$

$$\mathbf{M}_{t|t-1} = \mathbf{H}\boldsymbol{\Sigma}_{t|t-1} \quad (\text{A.7})$$

$$\mathbf{K}_t = \mathbf{M}'_{t|t-1}\mathbf{F}_{t|t-1}^{-1} \quad (\text{A.8})$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{F}_{t|t-1}\mathbf{K}'_t \quad (\text{A.9})$$

$$\mathbf{b}_{t|t} = \mathbf{b}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{y}_{t|t-1}) \quad (\text{A.10})$$

Conditional on the data \mathcal{Y}_T , parameters $\{\mathbf{G}, \mathbf{H}, \mathbf{Q}, \mathbf{N}\}$, and initial conditions $\{\mathbf{b}_{1|0}, \boldsymbol{\Sigma}_{1|0}\}$, the sequences of left hand side variables (A.3)-(A.10) are found by matrix multiplication.

\mathbf{Q} is a block diagonal matrix composed of \mathbf{V} , $(1/\zeta_1)^2$, $(1/\zeta_2)^2$, $\tilde{\mathbf{V}}$, and zeros otherwise. \mathbf{N} only has one nonzero entry: $N^{(2,2)} = (1/\zeta_\varepsilon)^2$. $\mathbf{b}_{1|0}$ is comprised of the true values of the revisions and economic variables, along with the estimate of the time-varying parameters from the model without data uncertainty for the starting time period of the model including data uncertainty.

$\boldsymbol{\Sigma}_{1|0}$ is a block diagonal matrix composed of $\mathbf{P}_{1|0}$ and the remaining elements specifying the initial uncertainty over the revisions and unemployment values. The off-diagonal elements are set to zero and the relevant diagonal elements are set to $\tilde{\mathbf{V}}^{(3,3)}$.

³²Technically, I must assume that the vector $\boldsymbol{\eta}_t$ appearing in (A.2) is Gaussian; assuming a Gaussian $\boldsymbol{\eta}_t$ for the general nonlinear case (3.8) does not assure that the shock in the first-order expansion would be Gaussian.

Likelihood The likelihood is

$$\mathcal{L}(\mathcal{Y}_T|\Psi) = \prod_{t=1}^T |\mathbf{F}_{t|t-1}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{y}_{t|t-1})' \mathbf{F}_{t|t-1}^{-1} (\mathbf{y}_t - \mathbf{y}_{t|t-1}) \right\}$$

where the $\mathbf{F}_{t|t-1}$ come from (A.6).

A.III Greenbook Forecasts

The main issue with comparing the model predicted unemployment forecasts to Greenbook forecasts is the difference in the frequency of observation. The model forecasts the monthly unemployment rate one month into the future. The Greenbook forecasts quarterly unemployment rates and are released without rigid frequency. For example, there are Greenbook forecasts published monthly through the 1970s, but into the 1980s and 1990s these forecasts are published almost at a bimonthly frequency. I take the following steps to make the comparison.

First, I form a quarterly unemployment rate series as the average of unemployment rate for the three underlying months. It is against these series that the forecasts produce forecast errors.

Second, I form a quarterly model-forecast series as the average of the step-ahead forecasts for the three underlying months. That is, the model's quarterly unemployment forecast for quarter q composed of months m_1, m_2, m_3 is

$$\frac{1}{3} (u_{m_1|m_1-1} + u_{m_2|m_1} + u_{m_3|m_2})$$

where $u_{j|j-1}$ is the forecast made at time $j-1$ pertaining to time j

Third, I form the quarterly Greenbook-forecast series as an average of all the forecasts made the month before or anytime during a quarter. That is, the Greenbook quarterly forecast for q composed of months m_1, m_2, m_3 and immediately preceded by month m_0 is

$$\frac{1}{n_{obs}} (gb_{m_0} + gb_{m_1} + gb_{m_2} + gb_{m_3})$$

where gb_j is the Greenbook forecast for quarter q published in month j . It should be noted that all four of these forecasts do not exist for every quarter, in which case only those observed are summed and n_{obs} adjusts to however many forecasts *are* observed.

The Diebold and Mariano (1995) statistic S_1 takes forecast error series $\{e_{it}\}$ and $\{e_{jt}\}$

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi \widehat{f_d(0)}}{T}}}$$

where

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T (e_{it} - e_{jt})$$

and I take $\widehat{f_d(0)}$ to be Andrews (1991) quadratic-spectral HAC estimator. The errors under consideration run from 1970 through 1995 so that $T = 104$. The forecast errors from the Greenbook and the two models are graphed in Figure A.1 (note different vertical scales).

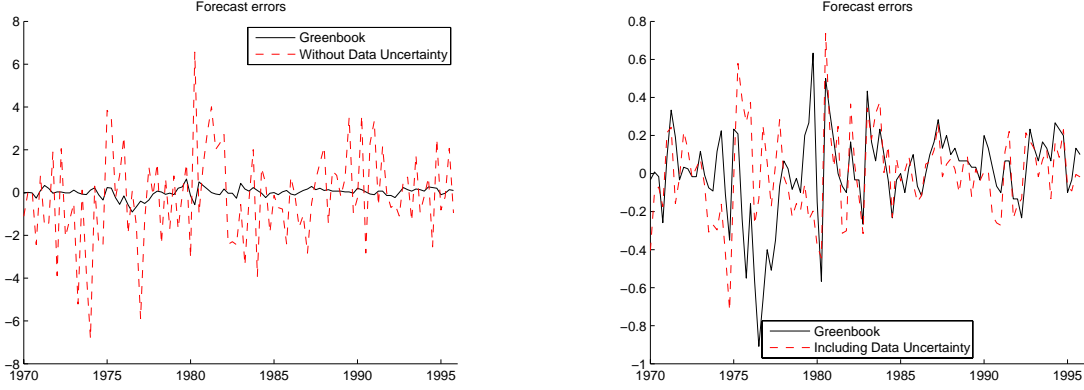


Figure A.1: FORECAST ERRORS ENTERING EQUAL-ACCURACY TEST.

Table A.1: REMAINING PARAMETER ESTIMATES

Model without Data Uncertainty – $\mathbf{P}_{1 0}$: standard deviations and <i>correlations</i>					
0.3273	0.9875	0.9187	-0.9953	-0.6626	-0.9959
	0.4412	0.9237	-0.9978	-0.6624	-0.9972
		0.0748	-0.9182	-0.9134	-0.9191
			0.7743	0.6545	0.9996
				0.0424	0.6515
					0.3111
Model including Data Uncertainty – $\mathbf{P}_{1 0}$: standard deviations and <i>correlations</i>					
1.573	-0.2426	-0.3637	-0.2986	-0.4698	0.3426
	0.4379	-0.6326	-0.8529	0.6846	-0.1535
		0.7839	0.8357	-0.5750	0.0095
			0.4790	-0.4403	-0.0264
				0.5759	-0.5757
					0.2482

Notes: The arrays are comprised as follows: the main diagonal is the square roots of the main diagonal of $\mathbf{P}_{1|0}$; the off-diagonal elements are the correlations derived from $\mathbf{P}_{1|0}$; $\mathbf{P}_{1|0}$ is the Fed’s initial step-ahead uncertainty over the initial Philips curve parameter estimate $\mathbf{a}_{1|0}$.

A.IV MCMC Implementation and Robustness

Priors The prior for Ψ is multivariate normal with a non-zero mean and a diagonal covariance matrix – so equivalently, the priors for each parameter are independent normals. The exact specifications are listed below where $\varphi \equiv (\text{vech}(\text{Chol}(\mathbf{V}))', \text{vech}(\text{Chol}(\mathbf{P}_{1|0})))'$ following the notation of Sargent, Williams, and Zha (2006):

$$\zeta_1 \mathcal{N}\left(5, \frac{5^2}{32}\right)$$

φ Follows Sargent, Williams, and Zha (2006). For each element on the diagonal of $\text{Chol}(\mathbf{V})$ or $\text{Chol}(\mathbf{P}_{1|0})$ the prior is $\mathcal{N}(0, 5^2 \times 0.5)$; for those elements off the diagonal, it is $\mathcal{N}(0, 2.5^2 \times 0.5)$

Convergence of the MCMC To address the convergence of the MCMC algorithm to its posterior distribution, I computed the number of iterations required to estimate the 0.025 quantile with a precision of 0.02 and probability level of 0.950 using the method of Raftery and Lewis

(1992). For each chain (with different initial conditions) the max of these across Ψ was below the $3.5E5$ iterations taken from each chain, suggesting that mixing the two chains (after burn-in) yields satisfactory precision.

Metropolis Algorithm An important part of the MCMC algorithm sampling from the posterior in a reasonable number of iterations is the covariance matrix of the proposal random step in the Metropolis algorithm. The Metropolis algorithm is

1. Given Ψ^{previous} , propose a new value

$$\Psi^{\text{proposal}} = \Psi^{\text{previous}} + \xi$$

where ξ is normal with mean zero and covariance matrix $c\Sigma_\xi$

2. Compute

$$q = \min \left\{ \frac{p(\Psi^{\text{proposal}}|\mathcal{Y}_T)}{p(\Psi^{\text{previous}}|\mathcal{Y}_T)}, 1 \right\}$$

3. Randomly draw $w \sim U(0, 1)$
4. If $w \leq q$, accept Ψ^{proposal} as current draw; otherwise, set Ψ^{previous} as the current draw

Given the manner in which all parameters affect the optimal policy, I arrived at this proposal covariance matrix Σ_ξ by doing the following. Using the covariance matrix for φ numerically solved for as described in Sargent, Williams, and Zha (2006)'s Appendix D and the prior covariance terms for all other elements of Ψ given above, the MCMC was started. For tens of thousands of iterations based on one initial condition, I considered only elements of the MCMC chain where a proposal had been accepted. From these chain elements I calculated the sample covariance matrix of the successful proposal shocks and set Σ_ξ equal to this. I tried different initial conditions and took the weighted average of the Cholesky factors of these sample covariance matrices. The tuning parameter c was adjusted to achieve an acceptance rate of around 25-35% during the first 20,000 iterations: after this, it was unadjusted, as continual chain-dependent adjustment of Metropolis step-size can negate the ergodicity upon which MCMC methods are based (see Robert and Casella (2004)).

Remaining $y_{t|t-1}$ forecasts Figure A.2 shows the remaining measurement equation predictions: current, 1-lag, 2-lag, and 72-lag unemployment report forecasts.

A.IV.1 Using Current Vintage CPI Inflation and Unemployment Rate Data

The plots in Figure A.3 are produced with inflation data as current vintage of CPI inflation (made available by those authors). The V and $P_{1|0}$ estimates are close to those reported in Table 1 for the model without data uncertainty while the estimate of ζ_1 is about one-half as large – they are available upon request.

A.IV.2 Using Preliminary Inflation and Unemployment Rate Data

The plots of Figure A.4 are produced by instead using preliminary CPI inflation and unemployment rate data. The V and $P_{1|0}$ estimates are close to those reported in Table 1 for the model without data uncertainty while the estimate of ζ_1 is about one-third as large – they are available upon request.

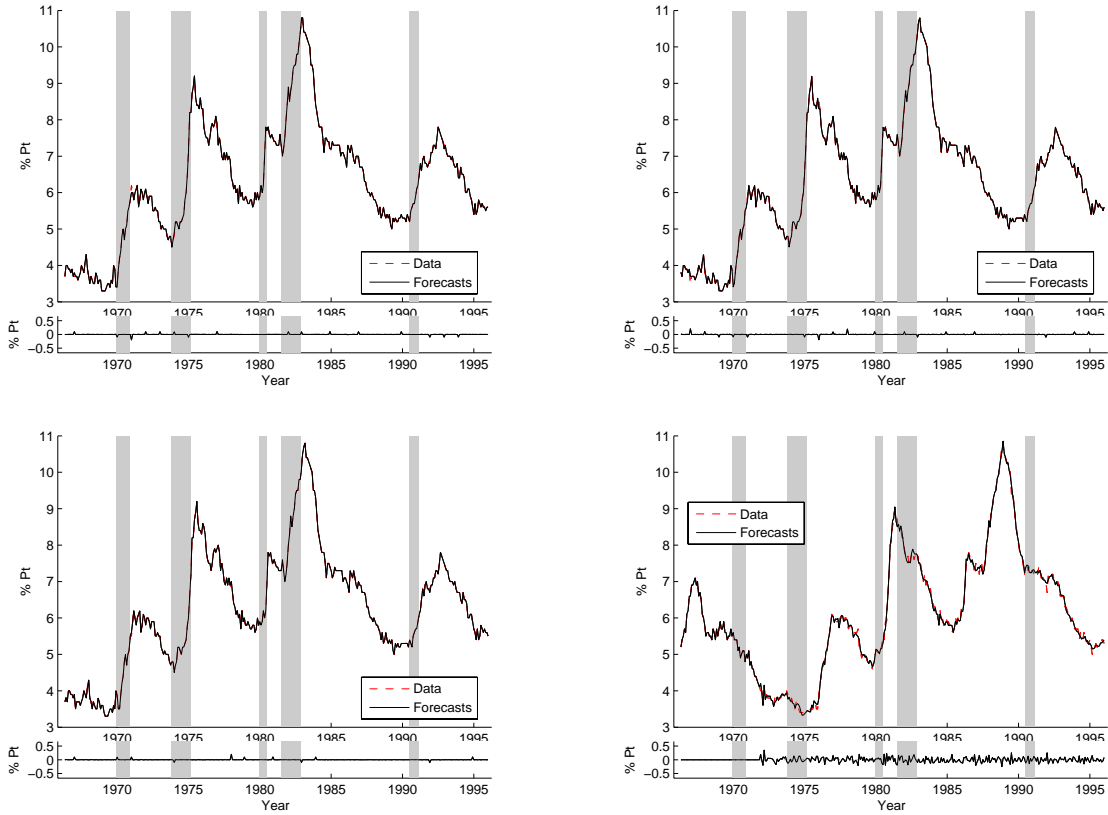


Figure A.2: FED UNEMPLOYMENT FORECASTS

Notes: now, one-step past-, two-step past-, and 72-step past-casts. NBER recessions are shaded. These figures are all on the same scale, but not on the scale of Figures 2 and 4.

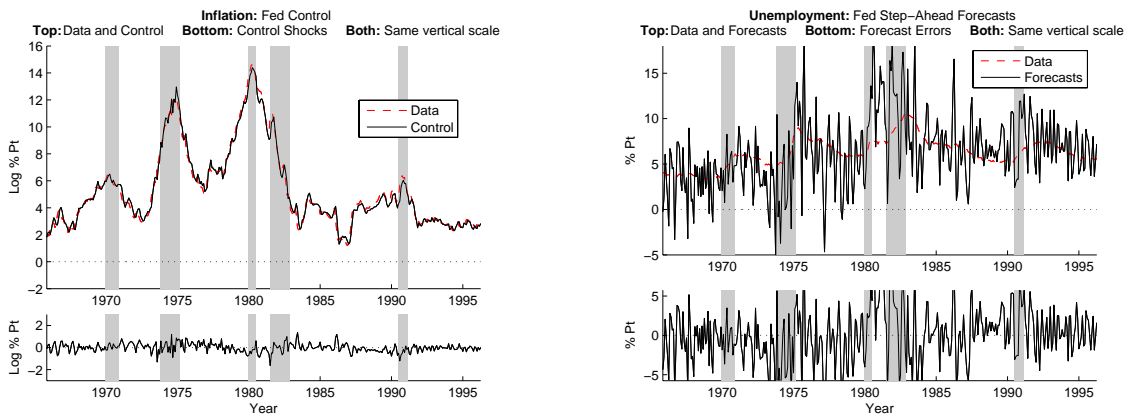


Figure A.3: FED INFLATION CONTROL AND UNEMPLOYMENT FORECASTS, CURRENT VINTAGE CPI

Notes: using current vintage seasonally-adjusted CPI inflation data. NBER recessions are shaded.

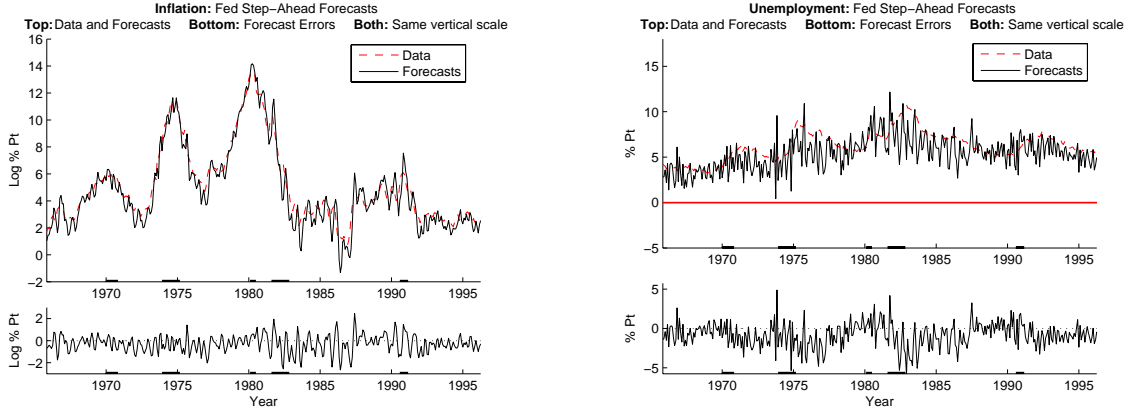


Figure A.4: FED INFLATION CONTROL AND UNEMPLOYMENT FORECASTS, PRELIMINARY CPI

Notes: using preliminary CPI inflation data (non-seasonally-adjusted). NBER recessions are shaded.

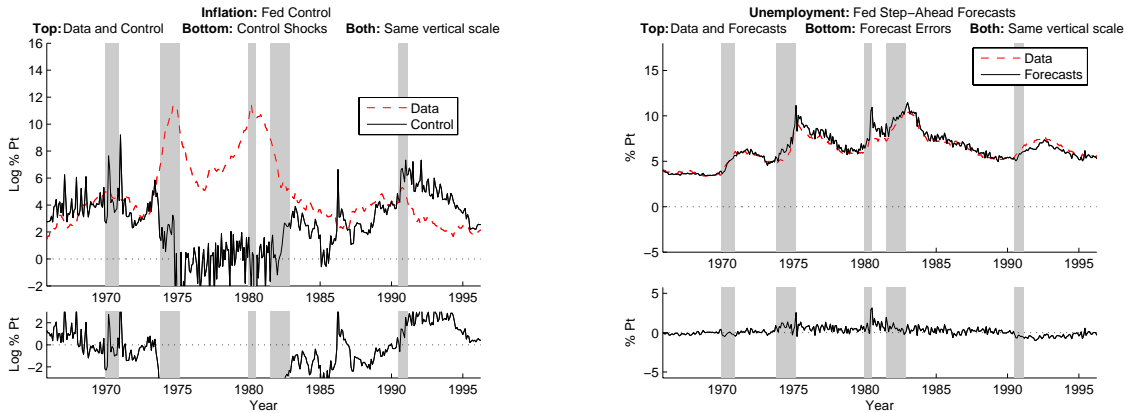


Figure A.5: FED INFLATION CONTROL AND UNEMPLOYMENT FORECASTS, MODIFIED LIKELIHOOD

Notes: using the model without data uncertainty and a modified likelihood placing more weight on the unemployment rate forecasts. NBER recessions are shaded.

A.IV.3 Modified Likelihood

It has been suggested on earlier drafts of this paper that the results may stem not so much from the data revisions as much as only the modified likelihood function taking account of revisions. This modification involves having the unemployment forecasts enter the likelihood, which might do most of the “smoothing” that is evident.

Doing this, we indeed see smoother Fed unemployment forecasts in the right panel of Figure A.5. As expected, the likelihood penalizes unemployment forecast errors and delivers far better ones. However, in order to accomplish this the time-varying Phillips curve estimates are such that inflation is far from target far too often, as the left panel shows.

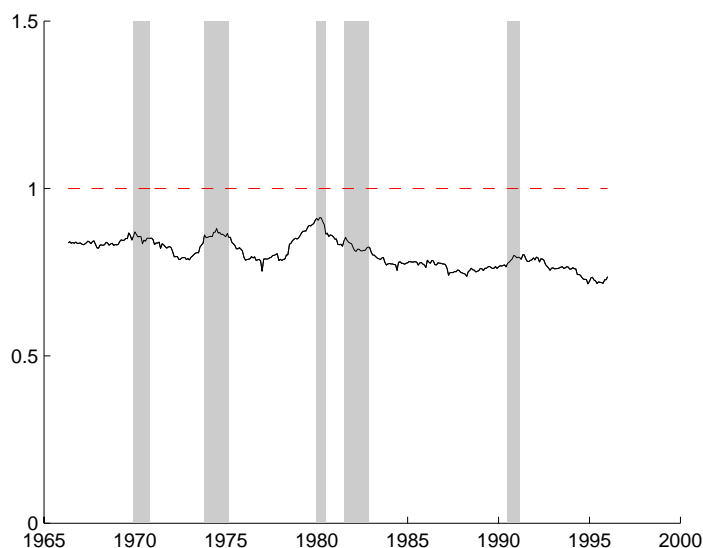


Figure A.6: MAXIMUM MODULUS EIGENVALUE, TIME-VARYING PARAMETER MODEL

Notes: maximum modulus of eigenvalue of characteristic polynomial of (3.2) coming from filtered time-varying parameter estimates in Section 3.3. NBER recessions are shaded.

A.V Autoregressive Stability

One issue that has been raised, for instance by Sims (2007), with the estimated beliefs coming from Sargent, Williams, and Zha (2006) is that they imply autoregressive instability for more than a dozen months around 1973. The model including data uncertainty addresses this “problem” – see Figure A.6. This section explains what exactly this issue is and also argues that it is not problematic to the purposes of using the Philips curve as a state transition equation.

Consider the autoregressive structure of the Philips curve the government is estimating over time. The estimates $\mathbf{a}_{t-1|t-1}$ determine the perceived transition law for the Fed’s optimal control problem.

Strictly-speaking, unemployment is not an explosive process: it must certainly take on values in $[0,100]$. Nonetheless, its high persistence, especially during the 1970s and 1980s, certainly makes it appear not “very” covariance-stationary. Consider estimating a bivariate VAR on the final data (f periods after the preliminary data) over rolling windows of varying sizes; Figure A.7 shows this for windows of 2, 3, 4, or 5 years.³³ If the Fed was concerned with breaks, a rolling window estimation procedure would be a straightforward way of picking this up. Notice that this eigenvalue condition is close to the unstable region most of the time. Hence, without positing a more sophisticated mechanism of Fed belief formation, we see that simple estimation methods would also have given evidence of unemployment instability.

Nonetheless, it is not clear that autoregressive instability would have led the Fed to reject

³³The window sizes are chosen with Friedman (1968) page 11 in mind: “[T]here is always a temporary trade-off between inflation and unemployment. . . I can at most venture a personal judgment, based on some examination of the historical evidence, that the initial effects of a higher and unanticipated rate of inflation last for something like two to five years.”

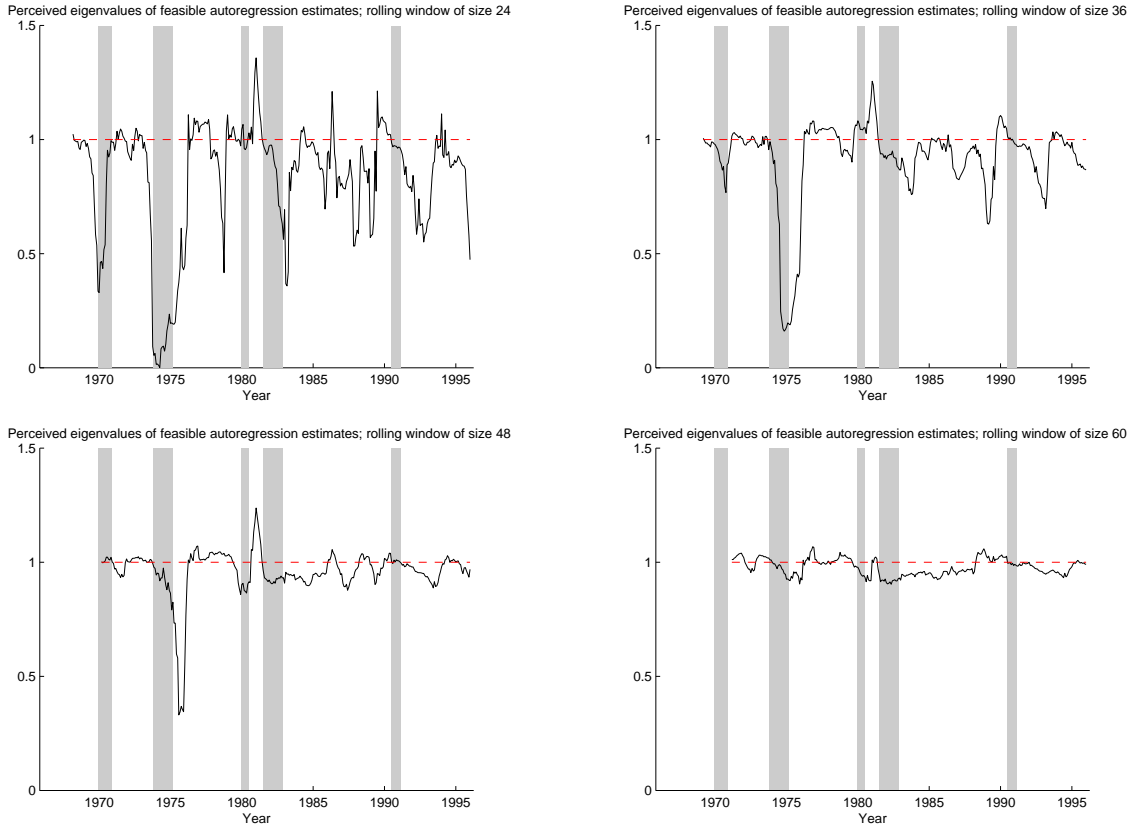


Figure A.7: MAXIMUM MODULUS EIGNEVALUE, BIVARIATE VAR

Maximum modulus of eigenvalue of characteristic polynomial of (3.2) coming from bivariate VAR; aligned so that the maximum modulus is at the time when it could be estimated with data at least f months ago. NBER recessions are shaded.

(3.2) as unemployment's dynamic structure for the purposes of setting optimal policy, given the accurate unemployment one-step-ahead predictions it yields (in the case where data uncertainty is acknowledged). To the optimal controller, what is important from forecasts are accuracy, not the description of the world they engender. The Fed had little reason to adjust its Philips curve just on the basis of these forecast errors. So the Fed may have done well to forecast unemployment using a rule which implied unemployment was not stationary, if the rule performed better.³⁴

³⁴For a related point, see the modeling motivation given in Primiceri (2005).