

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

Seth Pruitt[†]

August 2011

Abstract

Economic agents who are uncertain of their economic model learn, and this learning is sensitive to the presence of data uncertainty. I investigate this idea in a framework that successfully describes inflation as a learning Federal Reserve's optimal policy, but fails to satisfactorily motivate these policy shifts. I modify the framework to account for data uncertainty: The learning process is made more sluggish by its presence. Consequently the estimated model provides an explanation for the rise and fall in inflation: the concurrent rise and fall in the perceived Philips curve trade-off between inflation and unemployment.

Keywords: data uncertainty, extended Kalman filter, monetary policy, learning, markov-chain monte carlo, model uncertainty, optimal control, parameter uncertainty, real time data

JEL codes: E01, E58

*First version: February 2007. I thank Jim Hamilton for his insightful advice. Thanks also to the editor Ken West and anonymous referees for very helpful comments, as well as 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.

[†]Economist, Division of International Finance, Federal Reserve Board of Governors, 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 behavior had an important role in these events. (Clarida, Gali and Gertler 2000) and (Boivin 2006), among others, provide evidence of time-varying U.S. monetary policy responses over different parts of the postwar era. Some like (Romer and Romer 1990) and (Owyang and Ramey 2004) suggest that the changing response might be explained by changing Federal Reserve objectives. Alternatively, (Sargent 1999) suggests that evolving Federal Reserve beliefs about the economic environment can explain the rise and fall of inflation while holding the Federal Reserve's objectives constant.¹

This last explanation relies on the idea that agents learn about their economic environment, an idea notably advocated by (Evans and Honkapohja 2001). In such a framework, agents' prediction errors update their beliefs, represented as the parameters to their own economic model. These prediction errors are based on the data actually observed, which may contain measurement error. The point of this paper is that agents' data uncertainty, engendered by measurement errors, affects the evolution of their beliefs. Similar to (Brainard 1967), we see that this uncertainty tempers policy-makers' behavior: their learning is made more sluggish by data uncertainty.

To investigate the importance of this observation in practice, I modify the framework in (Sargent, Williams and Zha 2006) that itself extends (Sargent 1999). The Federal Reserve optimally controls inflation in light of constant unemployment and inflation targets, but is

¹(Orphanides and Williams 2006) and (Sims 2008) also argue that policy is sensitive to model uncertainty, and (Nason and Smith 2008) finds instability in the Philips curve relationship since the 1950s.

unconvinced that its economic model is correct and hence changes its beliefs in response to new data. Thus the framework attempts to explain the great American inflation as optimal policy caused by the Federal Reserve's changing estimates of a Philips curve linking inflation to unemployment. However, those existing results suffer from four problems. First, the Federal Reserve's unemployment rate forecasts, the basis for setting inflation, are inaccurate and much different than the Greenbook forecasts they should mimic. Second, the amount of estimated model uncertainty is very large, which undermines the plausibility of the story that the Federal Reserve believed in its estimated Philips curve enough to use it as the basis for policy. Third, the estimated beliefs imply that the Federal Reserve saw the unemployment rate as autoregressively explosive for various times in the 70s and 80s. Fourth, the framework explains the rise in inflation between 1970 and 1975, but does not give a good reason for the drastic fall in inflation between 1980 and 1985. Therefore, in these aspects the estimated model does not well motivate the economic story at the heart of the framework: inflation was raised in hopes of lowering unemployment until the Federal Reserve learned that such a trade-off was nonexistent.

Recently, (Carboni and Ellison 2009) modified (Sargent, Williams and Zha 2006)'s optimal control problem to target actual Greenbook forecasts, which successfully addresses the first two problems. However, an explanation for the rise *and fall* of inflation remains missing. The results here address all four problems, producing unemployment forecasts that are statistically indistinguishable from the Greenbook forecasts without using those Greenbook forecasts themselves as data.

I modify the optimal control framework by assuming the Federal Reserve recognizes that

data may be measured with error, the size of which I calibrate to evidence on data revisions. In line with this, and following the general suggestion of (Orphanides 2001), I differ from (Sargent, Williams and Zha 2006, Carboni and Ellison 2009) by fitting the model to real-time data. Data uncertainty introduces sluggishness into Federal Reserve’s learning process, and the ensuing framework is able to avoid the forementioned problems. The Federal Reserve’s model uncertainty drops significantly and the framework predicts unemployment forecasts that resemble Greenbook forecasts. Importantly, the results show a drop in the perceived Philips trade-off between 1980 and 1985 that explains the concomitant fall of inflation. 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.² What has been less clear is the impact this data uncertainty may hold for the purposes of economic research, and this paper suggests that data uncertainty can have a sizeable effect in frameworks where agents learn.

The paper is organized as follows. Section 2 introduces the main idea using a simple example. Section 3 introduces the frameworks – one without and the other including data uncertainty. Section 4 presents the estimation results for both frameworks, contrasts the results, and explains the factors leading to their differences. Section 5 concludes.

²In 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.

2 Why Data Uncertainty Matters in Learning Frameworks

This paper makes two simple points: One, the estimated size of modeled random shocks is positively biased by other sources of variation that are not modeled – there is omitted variable bias. Two, in a class of models with learning agents, their learning algorithm is nonlinear in the latent parameters and data, and so there is the possibility that small data measurement errors can be amplified into nontrivial amounts of bias. Here I explain exactly what is meant by these two statements.

Suppose an agent forms a forecast of an economic quantity y . The agent’s model maps data and parameters into this prediction. Parameter uncertainty is the situation where we allow past predictions and realized data to change the agent’s parameter vector going forward. For the purpose of this paper, we will think of models as completely described by particular parameter values. Hence, we will refer to this parameter uncertainty as *model uncertainty*.³ *Data uncertainty* is the situation where data is measured with error, possibly 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 the agent’s data uncertainty?

The intuition is straightforward that when the agent’s forecast is correct there is nothing to change about her model: The model’s performance cannot be improved.⁴ When the forecast error is nonzero, suppose the agent has an incentive to evaluate her model and learn from

³Because a collection of models may not be adequately indexed by parameter values, the concept of model uncertainty is more broad than that of parameter uncertainty. We restrict ourselves to the class of models where this indexation is sufficient, which is the one most often considered in the literature (Milani 2008, Eusepi and Preston 2010).

⁴This naturally follows from usual 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.

the error. At this point, the agent needs to understand *why* the forecast error is nonzero. To answer this, we consider the forecast error decomposition, which is the machinery that answers the question *why*.

Let $\boldsymbol{\epsilon}$ be a normal error associated with the agent-predicted parameter vector \mathbf{a} , $\boldsymbol{\varepsilon}$ be a normal error associated with the agent-predicted data vector \mathbf{x} , v be a normal error, all errors be mean zero, and each error be uncorrelated with the others. Suppose that the true data generating process is $y = (\mathbf{x} + \boldsymbol{\varepsilon})'(\mathbf{a} + \boldsymbol{\epsilon}) + v$. The agent knows this specification of the data generating process and makes a linear forecast $\hat{y} = \mathbf{x}'\mathbf{a}$. Finally, the agent knows that $\text{Var}(y - \hat{y})$ is given by

$$\sigma_v^2 + \mathbf{x}'\mathbf{P}\mathbf{x} + \mathbf{a}'\mathbf{Q}\mathbf{a} + c$$

where $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \mathbf{P}$, $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{Q}$, and $c > 0$, and these three quantities are known to the agent.⁵

However the researcher ignores the agent's data uncertainty, which means he assumes that $\boldsymbol{\varepsilon} = \mathbf{0}$. So the researcher decomposes the forecast error variance as $\tilde{\sigma}_v^2 + \mathbf{x}'\tilde{\mathbf{P}}\mathbf{x}$. But

$$\tilde{\sigma}_v^2 + \mathbf{x}'\tilde{\mathbf{P}}\mathbf{x} = \sigma_v^2 + \mathbf{x}'\mathbf{P}\mathbf{x} + \mathbf{a}'\mathbf{Q}\mathbf{a} + c,$$

and there will be a positive omitted variables bias to his estimates due to the presence of the omitted \mathbf{Q} and c . To clarify further, write the matrix $\begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 \end{pmatrix} = \delta\mathbf{D}$, assume that the researcher knows the true \mathbf{D} (determining the correlation structure and relative proportions of the variances), but must estimate the scalar δ (determining the average size

⁵The value c is the sum of expectations of fourth-order products of the elements of $\boldsymbol{\epsilon}$ and $\boldsymbol{\varepsilon}$. (Kan 2008, proposition 1) ensures that this c is positive, as well as that expectation of the terms involving v drop out.

of the parameter shocks). It is easy to see the researcher's estimate $\tilde{\delta}$ is biased:

$$\tilde{\delta} = \delta + \mathbf{a}'\mathbf{Q}\mathbf{a} + c.$$

This simple example captures what occurs in more complicated state space frameworks wherein agents learn. At each point in time the filtering algorithm delivers conditional means (the data and parameter predictions) and conditional covariance matrices (the errors' covariance structure). The forecast is a function of the means, and the forecast error is decomposed by a gain matrix that is a function of the covariance matrices. Obviously, the gain matrix can only decompose the observed forecast error to sources that are modeled, and so those sources are inferred to have taken on a larger realization than they actually did. This means that shocks to parameters are thought to be bigger than they actually are.

For a concrete example, time index and simplify the agent's model for y_t to involve only scalars, and then express it as an expansion around the point

$$(x_t, a_t, \varepsilon_t, \epsilon_t, v_t) = (x_{t|t-1}, a_{t|t-1}, 0, 0, 0).$$

Here we just take the values $x_{t|t-1}$ and $a_{t|t-1}$ as given (later they will be the output of the Kalman filter). The second-order expansion completely captures this function because the nonlinearity is very simple:

$$\begin{aligned} y_t &= x_{t|t-1}a_{t|t-1} + a_{t|t-1}(x_t - x_{t|t-1}) + x_{t|t-1}(a_t - a_{t|t-1}) + a_{t|t-1}\varepsilon_t + x_{t|t-1}\epsilon_t + v_t \\ &\quad (x_t - x_{t|t-1})(a_t - a_{t|t-1}) + (x_t - x_{t|t-1})\epsilon_t + (a_t - a_{t|t-1})\varepsilon_t + \epsilon_t\varepsilon_t \end{aligned} \quad (1)$$

and the linear predictor is $\hat{y} = x_{t|t-1}a_{t|t-1}$. Ignoring data uncertainty implies the assumption that $(x_t - x_{t|t-1})$ and ε_t are certainly zero. For simplicity, let's assume $\text{Var}(v_t) = 0$. Then

by ignoring data uncertainty we believe parameter shock ϵ_t to have the variance $\text{Var}(y - \hat{y})/2x_{t|t-1}^2$. But if there *are* measurement errors, that quantity is not $\text{Var}(\epsilon_t)$, but rather is $\text{Var}(\epsilon_t) + \frac{a_{t|t-1}^2}{x_{t|t-1}^2} \text{Var}(\varepsilon_t)$. Here we can see the bias that is introduced, and note it is a function of the measurement error and the squared levels of the parameters and data values. Therefore by ignoring data uncertainty we will believe that parameter shocks have taken larger values, on average, than they truly have.

The key idea is that the omitted variable bias is a function of the data error that may be amplified, depending on the values of parameters and data. It is a different matter to say whether or not this amplification happens in practice, and to investigate that possibility we turn to the next section.

3 Frameworks: With and Without Data Uncertainty

Why did inflation rise and fall so dramatically in the United States between 1970 and 1985? (Sargent, Williams and Zha 2006) reverse engineer Federal Reserve Philips curve beliefs that explain inflation as an optimal policy – they do so in a framework with model uncertainty but without data uncertainty. The Federal Reserve’s model evolves because we assume the Federal Reserve learns about the Philips curve. In light of the previous section, it is worth asking if the exclusion of data uncertainty significantly affects the framework’s predictions. We will see in Section 4 that the answer to this question is affirmative. But first we present the framework and modify it to explicitly account for Federal Reserve data uncertainty.

The Federal Reserve has a dual mandate to keep both the unemployment rate and the inflation rate low. Directly following (Sargent, Williams and Zha 2006), the Federal Reserve’s

objective function is written

$$\min_{\{x_{t-1+j}\}_{j=0}^{\infty}} \mathbb{E}_t \sum_{j=0}^{\infty} 0.9936^j \left((\pi_{t+j} - 2)^2 + (u_{t+j} - 1)^2 \right) \quad (2)$$

where x represents the Federal Reserve's setting (up to an exogenous shock) of inflation. I follow (Sargent, Williams and Zha 2006)'s choice of low inflation and unemployment rate targets, quadratic objective, and equal weighting of the two goals.

A quadratic objective function is assumed because it makes the problem very tractable and it is common in the literature. However, the symmetry of the loss function is troubling in principle because it says that if inflation/unemployment is *below* target, the Federal Reserve aims to *raise* them. By setting the targets to values that are almost always lower than has historically been seen for inflation and unemployment, we retain the tractability of the quadratic objective while delivering the message that the Federal Reserve has almost always tried to *lower* them, thus meeting the dual mandate. Since the dual mandate does not provide a weighting between the two goals, setting equal their relative weights seems reasonable.⁶

To achieve this objective the Federal Reserve controls the rate of inflation up to an exogenous additive shock. Therefore the annualized inflation rate π_t is

$$\pi_t = x_{t-1} + \zeta_1 \epsilon_{1t} \quad (3)$$

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

⁶Were we to lower the weight of the unemployment rate deviations, say, the model would level shift up the values of the Philips curve parameters presented below in order to match the data, but would not materially affect their time series properties.

To capture the Federal Reserve’s perceived relationship between inflation (which can be directly controlled up to an exogenous shock) and the unemployment rate (which cannot be directly controlled), we endow it with a model of a Philips curve in levels. However, we assume that the Federal Reserve is uncertain that its estimated model is correct, and so is constantly learning and updating its model estimate. We accomplish this by assuming the parameters follow a random walk, which introduces model uncertainty (and a motivation for learning) following the basic idea of (Cooley and Prescott 1976), expressed in the following two equations

$$u_t = \boldsymbol{\alpha}'_{t-1} \begin{pmatrix} \pi_t \\ \pi_{t-1} \\ u_{t-1} \\ \pi_{t-2} \\ u_{t-2} \\ 1 \end{pmatrix} + \zeta_2 \epsilon_{2t} \equiv \boldsymbol{\alpha}'_{t-1} \boldsymbol{\Phi}_t + \zeta_2 \epsilon_{2t} \quad (4)$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{Z}^{-\frac{1}{2}} \boldsymbol{\epsilon}_{3t}. \quad (5)$$

Because the parameters follow a random walk whose steps are independent of everything else, the Federal Reserve’s estimate of $\boldsymbol{\alpha}_{t-1}$ is also its estimate of $\boldsymbol{\alpha}_{t+j}$, $\forall j \geq 0$; thus we solve the problem for each period t using “anticipated utility”, following (Sargent 1999, Sargent, Williams and Zha 2006). Hence, the time t solution to the dynamic programming problem is found by using the Philips curve estimate $\boldsymbol{\alpha}_{t-1|t-1}$ as the law of motion for all $j \geq 0$; the time $t + 1$ plugs in the estimate $\boldsymbol{\alpha}_{t|t}$; and so forth.⁷

Finally, we assume that the data generating process for the unemployment rate does not respond to inflation in levels, but rather to inflation surprises: therefore the Federal Reserve’s model is drastically misspecified. We could follow (Sargent, Williams and Zha

⁷Since the objective function is quadratic, there is no impetus for the Federal Reserve to experiment. Anticipated utility follows (Marcet and Sargent 1989) and is an equilibrium concept often assumed in learning frameworks – for instance see (Eusepi and Preston 2010).

2006) and additionally estimate this process, but as (Carboni and Ellison 2009) point out this estimation is independent of the estimation of the Federal Reserve’s history of learning. Therefore we do not introduce this aspect of (Sargent, Williams and Zha 2006)’s model because it is immaterial to the focus of this investigation: the affect of data uncertainty on the estimated learning process.⁸

3.0.1 Without Data Uncertainty

If we assume the Federal Reserve ignores data measurement error, and therefore has no data uncertainty, then the Federal Reserve’s Philips curve estimates are the solution to a linear filtering problem where (5) is a state equation and (3)-(4) are observation equations:⁹

$$\boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\alpha}_{t|t-1} + \frac{\mathbf{P}_{t|t-1}\boldsymbol{\Phi}_t(\tilde{u}_t - \boldsymbol{\Phi}'_t\boldsymbol{\alpha}_{t|t-1})}{\zeta_2^2 + \boldsymbol{\Phi}'_t\mathbf{P}_{t|t-1}\boldsymbol{\Phi}_t}. \quad (6)$$

as in Sargent, Williams and Zha (2006). The information actually available to the Federal Reserve at time t is $\{\tilde{u}_t, \tilde{\pi}_t, \tilde{u}_{t-1}, \tilde{\pi}_{t-1}, \dots\}$, which are the available real-time data on inflation and unemployment rates (Orphanides 2001). When the Federal Reserve ignores data uncertainty, it assumes that the real-time data are the true values of the economic variables; that is, it assumes $\tilde{u}_t = u_t$ and $\tilde{\pi}_t = \pi_t$.

3.0.2 Including Data Uncertainty

On the other hand, there is evidence that data is subject to measurement error. Examples include: macroeconomic data revisions (Croushore and Stark 2001); data collection agen-

⁸This explains the exclusion of the “Lucas supply” equation from (Sargent, Williams, and Zha 2006) in this model exposition.

⁹The mean squared error evolution is given by

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} - \frac{\mathbf{P}_{t|t-1}\boldsymbol{\Phi}_t\boldsymbol{\Phi}'_t\mathbf{P}_{t|t-1}}{\zeta_2^2 + \boldsymbol{\Phi}'_t\mathbf{P}_{t|t-1}\boldsymbol{\Phi}_t} + \mathbf{Z}.$$

cies' documentation of collection errors; and, ample academic discussion on the intrinsic difficulty of definitive measurement of important macroeconomic concepts, such as potential output (Orphanides and van Norden 2002). This issue is common knowledge in policy and forecasting discussions (Cunningham, Jeffrey, Kapetanios and Labhard 2007, Pesaran and Timmermann 2005). It is reasonable to suppose a well-informed policy-maker like the Federal Reserve associates some uncertainty to the observed real-time data. In addition, Section 2's analysis suggested that omitted data uncertainty could increase the estimated volatility of an agent's beliefs. Foreshadowing the results to come, we will see that this is exactly the problem plaguing the framework without data uncertainty.

By accounting for data uncertainty, we make latent both the true values of economic variables and the true values of Federal Reserve's model parameters. Therefore (4) becomes a state equation and the state-space will be nonlinear in the state variables since latent data multiply latent parameters. This nonlinearity in the transition equation for the optimal control problem spreads also to the optimal policy rule equation delivering the Federal Reserve's inflation control variable x_{t-1} . Therefore, whereas previously (3) was a linear observation equation, here the relationship is placed in the state equation

$$\pi_t = x_{t-1}(\Phi_{t-1}) + \zeta_1 \epsilon_{1t}. \quad (7)$$

Here $x_{t-1}(\cdot)$ is the optimal policy that solves (2) subject to the (anticipated) law of motion (4) – this must be solved by numerical methods in each time period as the estimates of (4) evolve.

Due to the nonlinearity of the state equations, equations for the evolution of α are not given exactly by closed forms as above. Instead, we can collect the nonlinear equations

(4) and (7) together with the Philips curve evolution (5) into a function $\mathbf{g}(\boldsymbol{\beta}_t)$ for $\boldsymbol{\beta}_t \equiv (\boldsymbol{\alpha}'_t, u_t, \pi_t, u_{t-1}, \pi_{t-1}, u_{t-2}, \pi_{t-2})'$. We take the first-order expansion

$$\mathbf{g}(\boldsymbol{\beta}_{t-1}, \boldsymbol{\eta}_t) \approx \mathbf{g}_{t|t-1} + \mathbf{T}_{t|t-1}(\boldsymbol{\beta}_t - \mathbf{b}_{t-1|t-1}) + \mathbf{R}_{t|t-1}\boldsymbol{\eta}_t \quad (8)$$

where

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

This expansion defines the extended Kalman filter (Tanizaki 1996). Therefore, we must numerically calculate the derivative of x_{t-1} with respect to $\boldsymbol{\beta}_{t-1}$, and do so at each step in addition to finding the new optimal policy. Six of the rows in (8) now define the evolution of $\boldsymbol{\alpha}$, and it is apparent that they differ from (6) due to the presence of $\mathbf{T}_{t|t-1}$ and $\mathbf{R}_{t|t-1}$.¹⁰

Turning to the observation equations, we have

$$\tilde{\pi}_t = \pi_t + \zeta_4 \varepsilon_{4t} \quad (9)$$

$$\tilde{u}_t = u_t + \zeta_5 \varepsilon_{5t}. \quad (10)$$

In order to achieve identification apart from ζ_1, ζ_2 , the parameters of the data uncertainty process must be calibrated. To do this, I look at the revision between the first-reported value of the data and the most recent vintage of the data – see Section 3.1.1 below. I assume that

¹⁰Letting $\boldsymbol{\Sigma}$ be the mean squared error matrix around $\boldsymbol{\beta}$, its evolution is given by

$$\boldsymbol{\Sigma}_{t+1|t} = \mathbf{T}_{t+1|t} (\boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \mathbf{F}_{t|t-1} \mathbf{K}'_t) \mathbf{T}'_{t+1|t} + \mathbf{R}_{t+1|t} \mathbf{Q} \mathbf{R}'_{t+1|t}$$

for $\mathbf{K}_t = \mathbf{M}'_{t|t-1} \mathbf{F}_{t|t-1}^{-1}$, $\mathbf{M}_{t|t-1} = \mathbf{H} \boldsymbol{\Sigma}_{t|t-1}$ and $\mathbf{F}_{t|t-1} = \mathbf{H} \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}' + \mathbf{N}$ where \mathbf{H} is the observation equation matrix and \mathbf{N} is the covariance matrix of the observation equation errors.

the recent vintage is more accurate than the real-time data, and so define the measurement error as the observed revision.

This is a very simple method of capturing data uncertainty and it abstracts from explicitly modeling the data revision process. However, the results I report are robust to specifying a more detailed model for data revision. In particular, suppose that we included in the Federal Reserve’s observation vector not only these first-reported values of inflation and unemployment, but also revisions of past values. The complication this introduces is that the state space is expanded and new parameters are required to govern the revision processes possible serial correlation. Doing so, the estimates of the time-varying Federal Reserve beliefs are virtually unchanged. Why is this the case? Recall the simple example of Section 2: the important aspect is that we explicitly allow the Federal Reserve to treat the data itself as latent. Therefore, it will apportion some of its forecast error not only to parameter shifts but also the data measurement errors. This can easily be accomplished using (9) and (10).¹¹

3.1 Estimation

3.1.1 Data

Data on inflation (from headline CPI) and the civilian unemployment rate for ages 16 and older comes from the ALFRED, a real-time data archive established by the St. Louis Federal Reserve Bank: The frameworks use the first-reported values of these data series between 1960 and 1996. Inflation is calculated by the 12-month-ended logarithmic growth rate in the index. In making these choices, I balance two opposing goals: The aim to be as true as possible to the data the Federal Reserve would have actually had in hand, and the aim to as accurately

¹¹Further discussion is placed at the end of Section 4.2.

as possible present evidence on the amount of data uncertainty facing policy makers. The choice of headline CPI inflation and the unemployment rate achieves the first goal. Both of these are available in real-time over the entire sample. Figure 1 displays the time series of these revisions, defined as the difference between the most recent vintage value and the first-reported value. For both series the mean revision is small and the standard deviations are 0.106 for CPI inflation and 0.112 for the unemployment rate, leading to the calibration

$$\begin{pmatrix} \mu_4 \\ \mu_5 \end{pmatrix} = \begin{pmatrix} 0.005 \\ 0 \end{pmatrix}, \begin{pmatrix} \zeta_4^2 & 0 \\ 0 & \zeta_5^2 \end{pmatrix} = \begin{pmatrix} 0.106^2 & 0 \\ 0 & 0.112^2 \end{pmatrix}.$$

Moreover, in the case of CPI inflation, these revisions largely appear for (economically) unappealing reasons, like index re-basing of the index and seasonal adjustments.

However, there is evidence that revisions to CPI inflation and the unemployment rate give an inaccurate picture of the amount and nature of data uncertainty policy makers actually face. For example, the standard deviations of revisions to nonfarm payroll employment growth or GNP inflation are much larger – 0.39 and 0.54 percentage points, respectively – over the same period. Moreover, these other activity and inflation measures experience revisions due to economically reasonable factors like better measurement coming from low frequency data collections (like annual or quadrennial Census data). We take this to imply that generally policy-makers face a non-negligible amount of data uncertainty surrounding measurement of real activity and inflation. Nevertheless, this study solely uses only data on CPI inflation and the unemployment rate for its analysis, in the interest of parsimony: As such, the paper’s results could be interpreted as a lower bound on the effect of data uncertainty in such a learning framework.

3.1.2 Method

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). I have found little difference in practice between the first-order and second-order expansions and use the former to computational simplicity. This is due to the simple nature of the nonlinearity present in the key function $\alpha_t \Phi_t$. Again, recall the simple example in Section 2. The only difference between a first-order expansion and the second-order (1) are the cross-derivative terms on the second line. If the belief shock is uncorrelated with the data measurement error, these terms are zero on average. That lack of correlation appears to be a feature of the data, and so imposing the zeros in every period (by choosing the first-order expansion) makes little difference.

The parameter estimated is

$$\Psi \equiv \left(\zeta_1^{-1}, \zeta_2^{-1}, \text{vech}(\text{Chol}(\mathbf{Z}))', \text{vech}(\text{Chol}(\mathbf{P}_{1|0}))' \right)'$$

where $\text{Chol}(\cdot)$ is the Cholesky factor of positive definite matrix.¹² Because of Ψ 's large dimension, I follow (Sargent, Williams and Zha 2006) and use a Bayesian empirical method: a Markov-Chain Monte Carlo algorithm using the Metropolis-Hastings algorithm with random walk proposals to draw from the posterior distribution¹³

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

where $p(\Psi)$ is the prior and $\mathcal{L}(\mathcal{Y}_T|\Psi)$ is the likelihood (Robert and Casella 2004). From

¹²I estimate the reciprocals of ζ_1, ζ_2 because it is easy to draw them as normals and avoid nonnegativity constraints.

¹³I must use a accept/reject simulation technique because, due to the effects of Ψ on the whole sequence of forecasts, the form of (11) is not known.

the simulated posterior distribution I report medians as my point estimates and quantiles as probability intervals for the parameters. The initial value $\alpha_{1|0}$ is set to the regression estimate used by (Sargent, Williams and Zha 2006) based on non-real-time PCE inflation and unemployment data from January 1948 to December 1959, in order to maintain comparability to their results: the value is $\alpha_{1|0} = (-0.132, 0.141, 1.093, -0.022, -0.134, 0.219)$.

Table 1 displays the estimates for the framework without data uncertainty as the posterior median of 700,000 MCMC iterations from two separate runs of 400,000 with different initial conditions where the first 50,000 of each run is burned. The estimates in Table 1, and the following figures, are qualitatively similar to those of (Sargent, Williams and Zha 2006). However, they are not identical because here we use real-time data. To check MCMC diagnostics, 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 Raftery and Lewis' diagnostic – the estimates are well below 350,000 iterations per chain, indicated that we can be reasonable confident that the chains have run long enough.

The prior for Ψ is multivariate normal with a non-zero mean and a diagonal covariance matrix. That is, we have the priors: $\zeta_1^{-1}, \zeta_2^{-1} \mathcal{N}(5, 2)$; for each diagonal element of $\text{Chol}(\mathbf{Z})$ or $\text{Chol}(\mathbf{P}_{1|0})$ the prior is $\mathcal{N}(0, 5^2 \times 0.5)$; for each off-diagonal element, it is $\mathcal{N}(0, 2.5^2 \times 0.5)$.

For the framework including data uncertainty, I report the implied posterior distributions (approximated by a normal) obtained for the most important parameters: The diagonal elements of \mathbf{Z} and the ζ^{-1} s. ζ_1^{-1} has an approximately normal posterior distribution of about $\mathcal{N}(4.48, 0.21^2)$, ζ_2^{-1} an approximately normal posterior of about $\mathcal{N}(4.34, 0.76^2)$, and

the diagonal elements of \mathbf{Z} an approximately normal posterior with mean and variances

$$(0.011, 0.021, 0.057, 0.013, 0.048, 0.235), (0.12^2, 0.05^2, 0.03^2, 0.02^2, 0.14^2, 0.83^2).$$

These posteriors are somewhat close (in location) to the chosen priors, but have a tighter variance, indicating that the priors were chosen with the data in mind (they were, by looking at (Sargent, Williams and Zha 2006)'s results) but the likelihood adds some information as to their values. We can use the two-sample Kolomogorov-Smirnov test to compare the prior distribution to the posterior distribution. A multivariate distribution of this statistic is not easily available, so instead we compare the marginal distribution for each parameter. In all the cases considered, the test easily rejects: the “least” significant statistic is for ζ_2 , with a p -value below 0.001. We find this evidence to be suggestive of an informative likelihood surface, but not definitive since proper multivariate considerations are unavailable.

3.1.3 Identification

Note that we must calculate the derivative of the policy rule with respect to $\boldsymbol{\alpha}$ in order to find the Extended Kalman filter equation (8). This points to a distinction between the two frameworks described here. In the framework including data uncertainty ζ_2 is separately identified from \mathbf{Z} and $\mathbf{P}_{1|0}$. This identification follows from the fact that changes in \mathbf{Z} and $\mathbf{P}_{1|0}$ affect the first derivative of the optimal control policy rule while changes in ζ_2 have no such effect. Because the first derivative terms enter the expansion (recall (1)), the inflation equation (7) will respond differently to \mathbf{Z} and $\mathbf{P}_{1|0}$ than to ζ_2 , and therefore the likelihood will respond differently, and thus \mathbf{Z} and $\mathbf{P}_{1|0}$ are separately identified. Moreover, $\mathbf{P}_{1|0}$'s effect is distinct from \mathbf{V} 's effect since the former dies out quickly while the latter is identical for every period. Therefore, I impose that the value for ζ_2 in the framework without data

uncertainty is equal to the estimate from the framework including data uncertainty.¹⁴

On the other hand, without data uncertainty the parameters ζ_2 , \mathbf{Z} , and $\mathbf{P}_{1|0}$ can be scaled together with no effect on the likelihood – there is unidentification. (Sargent, Williams and Zha 2006) note this problem and overcome it by assuming that ζ_2 is one-tenth the size of the standard deviation of a “Lucas supply” equation for the unemployment rate that they additionally estimate.¹⁵ Because that equation has nothing to do with the Federal Reserve’s belief-generating mechanism, I refrain from specifying it at all, just as (Carboni and Ellison 2009) refrain during most of their paper.

4 Results

4.1 Estimates

Estimates of the frameworks’ parameters (perhaps more accurately referred to as hyperparameters) are contained in Table 1. With ζ_1 estimated at about $\frac{1}{5}$ in either model, the Federal Reserve believes that it has rather tight control of inflation. This is a feature shared by both frameworks, and we will see it corresponds to a close connection between actual inflation and the Federal Reserve’s model-implied optimal inflation policy. Since it is a major goal of this analysis, this result is quite important. Furthermore, the size of the additive Philips curve shock, ζ_2 , is around the same magnitude. Therefore in either model, we interpret the Federal Reserve as believing its Philips curve model *if it knew the right parameters* would

¹⁴For the sake of exposition, I place the estimates of $\mathbf{P}_{1|0}$ in the appendix because this parameter is far less important to the story and point of the paper than are \mathbf{Z} , ζ_1 , and ζ_2 .

¹⁵This has the effect of lowering the estimated values for \mathbf{Z} and $\mathbf{P}_{1|0}$. We will see later that high values of \mathbf{Z} imply a difficult economic story (namely that the Federal Reserve’s model of the economy was extremely unstable), and so to some extent the calibration of ζ_2 to *one-tenth* (as opposed to, say, *one-half* or *two times*) is not innocuous.

be an excellent basis for setting policy.¹⁶

However, not all estimates are so similar between the two frameworks. Turning to the estimates of \mathbf{Z} , the covariance matrix of the Philips curve parameter shocks, we see large and notable differences. In the framework without data uncertainty, notice that the estimated $Z^{(6,6)}$ implies that the Federal Reserve believes that every month the constant parameter is *extremely* volatile: the shock to this parameter has a standard deviation around 5 unemployment rate points, which appears rather implausible. Alternatively, this shock's standard deviation is around 25 basis points in the framework that accounts for data uncertainty.

In addition, in the framework without data uncertainty there appears to be a tight connection between the shocks to the Philips curve constant and the shocks to the parameters on lagged unemployment and inflation. Alternatively, in the framework including data uncertainty this shock is relatively unrelated to the other parameters' shocks. How are these values borne out by the estimated series of Federal Reserve optimal policy and unemployment rate forecasts?

Turning to Figure 2, we see that the small values of ζ_1 are connected to an estimated policy that closely tracks the actual history of American inflation. Indeed, the model is written down to do exactly this, and so it may not be surprising that it succeeds. But, as do (Sargent, Williams, and Zha 2006, Carboni and Ellison 2009), it is worth noting that this prediction of the model is very promising.

What is left to do is to check the remaining aspects of the economic narrative that this framework embodies. The core story is that inflation is an optimal policy implied by the

¹⁶Recall that in the model without data uncertainty ζ_2 is not estimated, but instead set to the value estimated in the model including model uncertainty.

Federal Reserve’s evolving model of the Philips curve trade-off between inflation and unemployment. We therefore focus on three key aspects of this evolving model: the closeness of the model-implied *unemployment rate forecasts* to actual Federal Reserve unemployment rate forecast data from historical Greenbooks, the *economic implications* of the model-implied economy at any point in time, and the strength of the predicted *economic motivation* for raising and lowering inflation.

4.1.1 The Federal Reserve’s Evolving Model: Unemployment Rate Forecasts

An important aspect of the model-predicted Federal Reserve beliefs are what they deliver in terms of unemployment forecasts: these are plotted in Figure 3. For the framework without data uncertainty, these forecast errors have a mean of 0.200 percentage points and a standard deviation of 1.904 percentage points. This slight downward bias of the unemployment rate forecasts is reasonable given the main point of the model: inflation is raised because it is forecasted to lower unemployment. But it seems difficult to believe that the Federal Reserve would persist in using a model whose forecast errors on average are almost twenty times larger than the usual change in the unemployment rate.¹⁷

In the bottom panel we plot the unemployment rate forecasts from framework including data uncertainty, and there we see a remarkable difference. The Federal Reserve’s forecasts are considerably more accurate than before: the forecast error standard deviation is 0.250 percentage points and the mean is basically zero. That is, by including data uncertainty, we estimate that the Federal Reserve’s Philips curve was a rather effective forecasting instrument. It is reasonable that such a model, though we now know it was drastically misspecified,

¹⁷The usual change would be the standard deviation of the forecast error in a random walk model.

might for a time be believed enough to set inflation policy.

Even more convincing is a direct comparison between these model-implied unemployment rate forecasts, and actual data on historical Federal Reserve unemployment rate forecasts contained in Greenbooks.¹⁸ A simple test of similarity between these forecasts is the DMW statistic of of (Diebold and Mariano 1995, West 1996).¹⁹ The DMW statistic tests the null hypothesis that the predicted unemployment forecasts have accuracy equal to the Greenbook forecasts. Testing the similarity between Greenbook forecasts and forecasts from the framework without data uncertainty results in an absolute value of the DMW statistic of 8.071 whose p -value is less than 0.001. Therefore we can say that the framework without data uncertainty produces forecasts that are statistically different from the Greenbook forecasts they should resemble. Comparing the Greenbook forecasts to the forecasts from the framework including data uncertainty, the DMW statistic takes the value 0.872 with p -value equal to 0.383. Therefore, we accept the hypothesis that this framework predicts Federal Reserve unemployment rate forecasts are similar to actual Greenbook forecasts in terms of statistical accuracy.

As might be expected, there is a statistical difference between the two frameworks' predicted forecasts: the DMW statistic takes the absolute value of 7.767 with p -value less than 0.001. Moreover, suppose we run the regression of Greenbook forecast errors on the forecast errors coming from each framework. The coefficient on the errors from the framework without data uncertainty is negative and insignificant, while the coefficient on the errors from the framework including data uncertainty is 0.600 with t -stat of 3.104. Thus, there

¹⁸It is exactly this point that prompts (Carboni and Ellison 2009)'s modification to (Sargent, Williams and Zha 2006)'s basic framework.

¹⁹The appendix discusses these forecasts and provides more details on the test statistic.

is evidence that the latter errors helps explain the Greenbook errors, while the former errors do not.²⁰ It should be noted that this has been achieved without actually using those Greenbook forecasts as data.

4.1.2 The Federal Reserve’s Evolving Model: Economic Implications

Each estimated framework delivers different implications about the historical evolution of the Federal Reserve’s Philips curve model. We focus on what this model would have told the Federal Reserve about the unemployment rate process. Two points emerge: One, a problem facing the framework without data uncertainty is that the estimated Philips curve is at some points autoregressively unstable. Two, by excluding data uncertainty the estimate of the Philips curve constant displays a great deal of volatility and movement – later we’ll refer to this evolution in our discussion of what drives the differences between these two frameworks.

The first critique involves the autoregressive character of the unemployment rate that the estimated Philips curve implies along its evolution. The Philips curve describes current unemployment in terms of inflation and lags of unemployment – this defines an autoregression for unemployment. So what is the character of unemployment’s dynamic relationship as estimated at various points in time? This can be summarized by the roots of the Philips-curve-implied characteristic polynomial, or equivalently the maximum (in modulus) eigenvalue of the Philips-curve-implied “companion-form” VAR(1) matrix.²¹ In the framework without data uncertainty, this maximum eigenvalue rises above 1 during the mid-1970s, a point made by Christopher Sims in his discussion of the draft version of (Sargent, Williams,

²⁰Nonetheless, the coefficient is 0.6 not 1, and the R^2 is 0.095: there is more going on in Greenbook forecasts than this simple model can capture.

²¹ \mathbf{F} in (Hamilton 1994, p.259).

and Zha 2006). Thus the model predicts that during these times the Federal Reserve regarded the unemployment rate as dynamically unstable, not stationary. On the other hand, in the framework including data uncertainty this maximum modulus eigenvalue stays well below unity for all periods.

The second point is that the constant in the Philips curve follows dramatically different evolutions, depending on whether or not data uncertainty is included to affect its estimates. Figure 4 plots the evolutions from the two frameworks. In the framework without data uncertainty, the rise of inflation happens at the same time as this constant jumps. This is the model trying to reconcile the growing unemployment rate during 1970–1975 with (i) a jump in the perceived Philips curve trade-off (more on this below), and (ii) rising inflation. The Philips curve model suggested at this time that the effect of the inflation level was increasingly expansionary, but the unemployment data continued to come in weak: therefore, the estimate of the Philips curve constant, a main driver of the perceived natural rate of unemployment, jumps. One can see from the left panel of Figure 4 that this constant parameter is very volatile. Such large shifts in its Philips curve would have been early and persuasive evidence for the Federal Reserve to reconsider setting policy according to this model.

On the other hand, in the framework including data uncertainty this parameter’s evolution is much more gradual. In many ways, it traces out the pattern of actual inflation, because it is being relied upon to explain the low frequency movement in the unemployment rate that do not make sense according to the Federal Reserve’s perceived Philips curve trade-off until the latter falls.²² So what of this trade-off?

²²It should be noted that the implied natural rate of unemployment reaches implausible levels in both

4.1.3 The Federal Reserve’s Evolving Model: Economic Motivation

The Federal Reserve’s beliefs about the Philips curve trade-off are the economic heart of this simple framework. It must provide the motivation for why policy would have set inflation high above target: recalling the Introduction, maintaining constant Federal Reserve objectives, the reason for taking on painful inflation is to encourage employment. Therefore, the evolution of this Philips curve trade-off is of primary importance to the thrust of this narrative. The frameworks’ trade-offs are plotted in Figure 5: to focus in on the relevant information, we plot both on the same scale between 1970 and 1985.

In the top panel, the trade-off in the framework without data uncertainty experiences a large jump between 1970 and 1975 which explains the great rise in inflation over those years. Thereafter the trade-off stays high, with no sharp activity around the disinflation of the early 1980s. 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 framework without data uncertainty, the dramatic fall of inflation between 1980 and 1985 occurs without any sharp change in the Federal Reserve’s beliefs.²³

On the other hand, consider the trade-off in the framework including data uncertainty, depicted in the bottom panel of Figure 5. Here the we predict a steady drop in the Philips curve trade-off starting around 1980. As this perceived trade-off diminishes, inflation falls by more than 80% off its peak going into 1985.

estimated frameworks: 70% in the framework without data uncertainty, and 30% in the framework including data uncertainty. This appears to be a weakness of such a simple model. These results are available from the author upon request.

²³Estimated trade-offs in (Sargent, Williams and Zha 2006, Carboni and Ellsion 2009) are similar to the one shown here for the framework without data uncertainty.

Thus, the framework including data uncertainty describes a consistent connection between the Federal Reserve’s inflation control and the Federal Reserve’s beliefs about the Philips curve trade-off: The Federal Reserve had a misspecified Philips curve model whose best estimate adapted to in-coming data and determined optimal policy. This policy caused the great American inflation in hopes of bringing down unemployment, until the Federal Reserve learned that such a trade-off was nonexistent.

4.2 How Data Uncertainty Makes a Difference

How can such a small amount of data uncertainty change the results? There are two parts to the answer. First: The evident data uncertainty is actually a fair bit larger than the Federal Reserve’s model uncertainty (particularly the variances of the parameters on inflation and unemployment in the Philips curve) to begin with. This is visible from the bottom panel of Table 1, where most of the parameter shocks’ standard deviations are only a fraction of the measurement error standard deviations that are calibrated to around 0.10 for both inflation and the unemployment rate. Second: Introducing measurement error into the learning framework creates a nonlinearity (recall Section 2) that amplifies the modest amount of data uncertainty into a biased estimate of model uncertainty.

This bias is important for decomposing the forecast error into the part contributed by parameter shocks, and varies over time because it depends on the value of each period’s data and parameter predictions. During the sample period, this bias is always greater than 10% and attains values around 150% – this means that the average parameter shock is perceived to be 1.1–2.5 times larger than it actually was.

Therefore, by ignoring the Federal Reserve’s data uncertainty, we overestimate the size

of the shocks to the Philips curve coefficients on inflation and unemployment. Accordingly, these parameters shift a lot from period to period in response to the large unemployment rate forecast errors; thus, there are shifts in the optimal policy rule's dependence on past inflation and unemployment. But inflation over this period is rather persistent: so each period *ceteris paribus* the current optimal policy should be somewhat near last period's policy in order to fit the data. Hence, the constant in the Philips curve adjusts to offset the change in the policy rule engendered by the shifting Philips curve coefficients on inflation and unemployment – recall Figure 4. In turn, this makes the model consistent with a great deal correlation between the constant and the other parameters in the Philips curve, as is visible in the last column of the top panel of Table 1.

On the other hand, by explicitly taking account of data uncertainty, we estimate a lower correlation between the constant and other Philips curve parameters, as is visible in the last column of the bottom panel of Table 1. The key is introducing measurement errors which make past inflation and unemployment latent. The Federal Reserve is more sluggish to change its Philips curve estimate on the basis of the observed forecast errors, inferring that some of this forecast error may be due to error in measuring past values of inflation and unemployment on which the forecast is based. By explicitly accounting for the data uncertainty, we obtain an intuitive message that is somewhat familiar. In (Brainard 1967) the point was made that a policy-maker's model uncertainty leads to an optimal *dampening* of policy. Here the point is that a policy-maker's data uncertainty leads to an optimal *dampening* of learning.

These results highlight two points in favor of explicitly including data uncertainty in

economic frameworks that specify agents' model uncertainty. First, small amounts of data uncertainty can be amplified by the learning process. As we have just seen, an economic researcher may make substantially different inferences depending on whether or not data uncertainty is accounted for, even if the amount of uncertainty they would ignore is small. Second, we are able to discipline this modeling feature with the facts. We have evidence, one form of which are data revisions, to point to when discussing data uncertainty and ascertaining how large it may be. The role of data uncertainty can be disciplined by observables, and this makes it a potentially constructive part of learning frameworks.

4.2.1 Sensitivity

This paper has taken the most parsimonious route to introducing and calibrating data uncertainty. What is the sensitivity of these results to the calibration chosen? And what would happen if a more realistic data revision process were modeled?

With regard to calibration, recall that the characterization of the measurement error is taken straight from the data. The nonzero mean is completely innocuous. Were it changed to zero, the filter would additively adjust its predictions to take account of this, and this adjustment would be slight. The important aspect of the calibration is the standard deviation of the measurement error. If the standard deviations are lowered, we see the estimated framework start to resemble the framework without data uncertainty: this happens rather quickly at values below one-twentieth of a percent. If the standard deviations are raised, we see the estimated framework deliver unemployment rate forecasts that are similar to the framework without data uncertainty, starting at values of the unemployment rate measurement error standard deviation around half a percent. The estimated Philips curve trade-off

evolution flattens out quite a bit, and the Philips curve constant is called upon to do the lion's share of matching policy to actual inflation – but this leads to wild unemployment rate forecasts. Therefore, there is something like a sweet spot for data uncertainty in this particular application: fortunately the evidence suggests a value right in the middle of this region.

A more realistic data revision process is certainly of interest. In fact, earlier drafts of this paper included detailed modeling of this process, complete with nontrivial correlation structures between revisions of different horizons. However, the extra machinery is unnecessary – the crucial ingredient was found to be the unconditional standard deviations of the unemployment and inflation rate measurements actually entering the Philips curve. This is because the filter uses these parameters to determine the gain matrix, which is the key to decomposing the forecast error into estimates of the parameter and data measurement shocks. The entire intuition is contained in the simple example of Section 2 and revolves around omitted variable bias. In this application, one can be more realistic in calculating the appropriate gain matrix by modeling the data revision process in detail; but a simpler, and just as effective, method is to specify data measurement error as this paper has done.

5 Conclusion

This paper analyzes the effects of data uncertainty in frameworks with agents who learn. Ignoring data uncertainty can bias estimates of agents' model uncertainty. I investigate the importance of this point by extending the framework of (Sargent, Williams and Zha 2006) and showing that the explicit modeling of data uncertainty remedies some well-known issues

with that paper's results. The mechanism by which data uncertainty matters is through introducing sluggishness into the Federal Reserve's learning. Once this is the case, the framework predicts that the inflation of the 1970s and 1980s can be explained by evolving beliefs about the Philips curve trade-off between inflation and unemployment.

References

- [1] Brian D O Anderson and John B Moore. *Optimal Filtering*. Prentice-Hall, 1979.
- [2] Donald W K Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–58, May 1991.
- [3] Jean Boivin. Has us monetary policy changed? Evidence from drifting coefficients and real-time data. *Journal of Money, Credit, and Banking*, 38(5):1149–1173, August 2006.
- [4] William Brainard. Uncertainty and the effectiveness of policy. *The American Economic Review: Papers and Proceedings*, 57:411–425, 1967.
- [5] Giacomo Carboni and Martin Ellison. The great inflation and the greenbook. *Journal of Monetary Economics*, 56(6):831–841, September 2009.
- [6] Richard Clarida, Jordi Gali, and Mark Gertler. Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics*, 115(1):147–180, February 2000.
- [7] Thomas F Cooley and Edward C Prescott. Estimation in the presence of stochastic parameter variation. *Econometrica*, 44(1):167–84, January 1976.
- [8] Dean Croushore and Tom Stark. A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130, 2001.
- [9] Alastair Cunningham, Chris Jeffery, George Kapetanios, and Vincent Labhard. 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, February 2007.
- [10] Francis X Diebold and Roberto S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63, July 1995.
- [11] Graham Elliott, Ivana Komunjer, and Allan Timmermann. Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, 72(4):1107–1125, October 2005.
- [12] Stefano Eusepi and Bruce Preston. Central bank communication and expectations stabilization. *American Economic Journal: Macroeconomics*, 2(3):235–71, July 2010.
- [13] George W. Evans and Seppo Honkapohja. *Learning and Expectations in Macroeconomics*. Princeton University Press, 2001.
- [14] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [15] Raymond Kan. From moments of sum to moments of product. *Journal of Multivariate Analysis*, 99(3):542–554, March 2008.
- [16] Albert Marcet and Thomas J Sargent. Convergence of least-squares learning in environments with hidden state variables and private information. *Journal of Political Economy*, 97(6):1306–22, December 1989.
- [17] Fabio Milani. Learning, monetary policy rules, and macroeconomic stability. *Journal of Economic Dynamics and Control*, 32(10):3148–3165, October 2008.

- [18] James M. Nason and Gregor W. Smith. The new keynesian phillips curve : lessons from single-equation econometric estimation. *Economic Quarterly*, (Fall):361–395, 2008.
- [19] Athanasios Orphanides. Monetary policy rules based on real-time data. *American Economic Review*, 91(4):964–985, September 2001.
- [20] Athanasios Orphanides and Simon van Norden. The unreliability of output-gap estimates in real time. *Review of Economics and Statistics*, 84(4):569–583, Nov 2002.
- [21] Athanasios Orphanides and John C. Williams. Monetary policy with imperfect knowledge. *Journal of the European Economic Association*, 4(2-3):366–375, 04-05 2006.
- [22] Michael T. Owyang and Garey Ramey. Regime switching and monetary policy measurement. *Journal of Monetary Economics*, 51(8):1577–1597, November 2004.
- [23] Hashem Pesaran and Allan Timmermann. Real-time econometrics. *Econometric Theory*, 21:212–231, 2005.
- [24] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2 edition, 2004.
- [25] Christina D. Romer and David H. Romer. Does monetary policy matter? a new test in the spirit of friedman and schwartz. NBER Working Papers 2966, National Bureau of Economic Research, Inc, February 1990.
- [26] Thomas J. Sargent. *The Conquest of American Inflation*. Princeton University Press, 1999.

- [27] Thomas J. Sargent, Noah Williams, and Tao Zha. Shocks and government beliefs: The rise and fall of american inflation. *American Economic Review*, 96(4):1193–1224, September 2006.
- [28] Christopher A. Sims. Improving monetary policy models. *Journal of Economic Dynamics and Control*, 32(8):2460–2475, August 2008.
- [29] Hisashi Tanizaki. *Nonlinear Filters: Estimation and Applications*. Springer-Verlag, 2 edition, 1996.
- [30] Kenneth D West. Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–84, September 1996.
- [31] Arnold Zellner. 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, March 1958.

A Appendix

The estimates of $P_{1|0}$ from both models are reported in Table 2.

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, West 1996) statistic DMW takes forecast error series $\{e_{it}\}$ and $\{e_{jt}\}$

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

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)'s quadratic-spectral HAC estimator. The errors under consideration run from 1970 through 1995 so that $T = 104$, and I compare DMW to a standard normal distribution.

Table 1: PARAMETER ESTIMATES

Without Data Uncertainty						
ζ_1 :	0.213	(0.196,0.225)				
<i>Z</i> : standard deviations and <i>correlations</i>						
0.296	-0.985	0.154	0.884	-0.109	-0.285	
	0.279	-0.076	-0.879	0.168	0.392	
		0.161	0.382	0.801	0.795	
			0.190	0.292	-0.039	
				0.322	0.747	
					5.079	
Including Data Uncertainty						
ζ_1 :	0.223	(0.195,0.254)				
ζ_2 :	0.230	(0.205,0.265)				
<i>Z</i> : standard deviations and <i>correlations</i>						
0.011	0.967	-0.975	-0.823	0.952	-0.123	
	0.021	-0.989	-0.940	0.954	0.001	
		0.057	0.895	-0.987	0.125	
			0.013	-0.833	-0.235	
				0.048	-0.269	
					0.235	

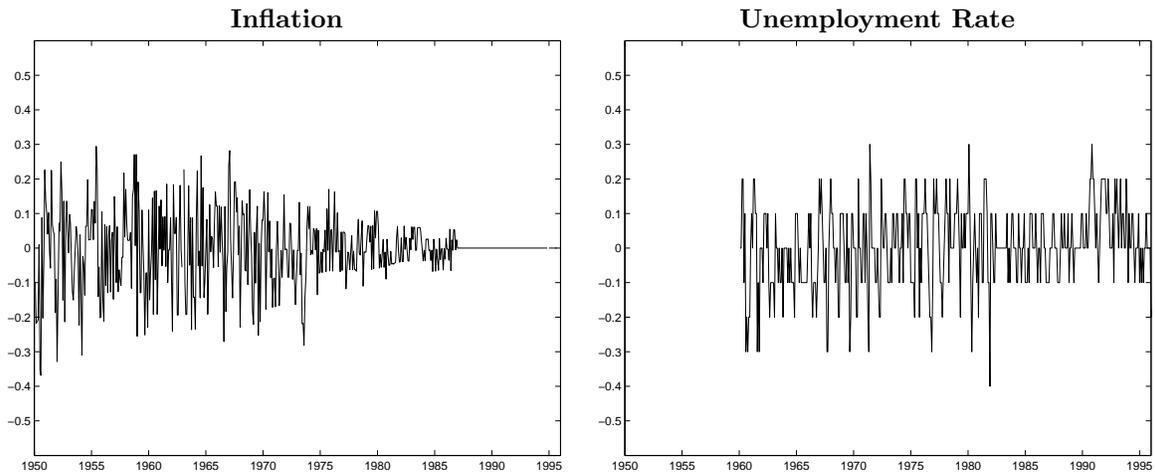
Notes: median of posterior distribution. ζ_1 is the standard deviation of ϵ_{1t} , the additive shock to the Federal Reserve's inflation control; ζ_2 is the standard deviation of the additive shock in the Federal Reserve's Philips curve. 95% probability intervals in parentheses. The bottom array is comprised as follows: the main diagonal are the square roots of the main diagonal of \mathbf{Z} ; the off-diagonal elements are the correlations derived from \mathbf{Z} ; \mathbf{Z} is the covariance matrix of the ϵ_{3t} shock to the time-varying Philips curve 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 .

Table 2: PARAMETER ESTIMATES – APPENDIX

Without Data Uncertainty					
$\mathbf{P}_{1 0}$: standard deviations and <i>correlations</i>					
0.327	0.988	0.919	-0.995	-0.663	-0.996
	0.441	0.924	-0.998	-0.662	-0.997
		0.075	-0.918	-0.913	-0.919
			0.774	0.654	0.999
				0.042	0.652
					0.311
Including Data Uncertainty					
$\mathbf{P}_{1 0}$: standard deviations and <i>correlations</i>					
1.573	-0.243	-0.364	-0.299	-0.469	0.343
	0.438	-0.633	-0.853	0.685	-0.154
		0.784	0.836	-0.575	0.009
			0.479	-0.440	-0.026
				0.576	-0.576
					0.248

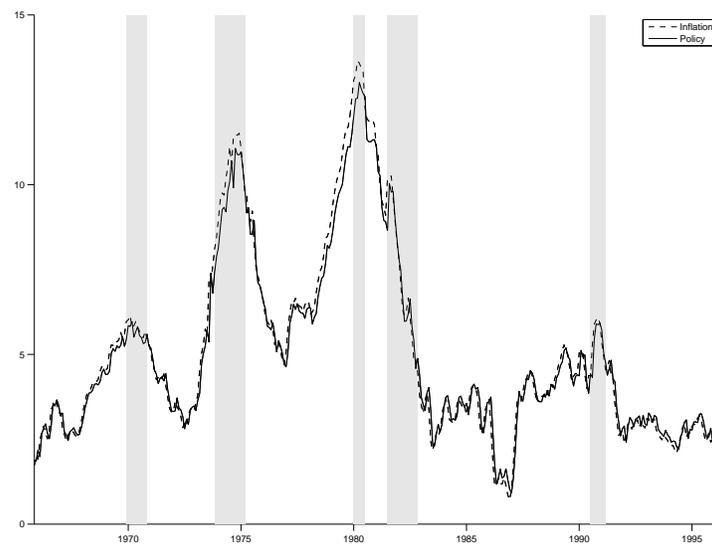
Notes: The arrays are comprised as follows: the main diagonal is the square roots of the main diagonal of $\mathbf{P}_{1|0}$; the off-diagonal elements are the correlations derived from $\mathbf{P}_{1|0}$; $\mathbf{P}_{1|0}$ is the Federal Reserve's initial step-ahead uncertainty over the initial Philips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

Figure 1: DATA REVISIONS

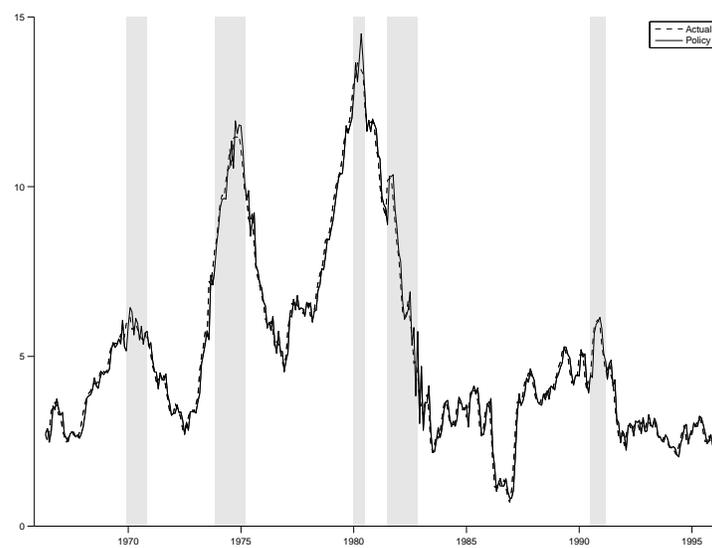


Notes: Defined as difference between the most recent vintage value and the first-reported value.

Figure 2: INFLATION RATE VS. OPTIMAL POLICY
Without Data Uncertainty

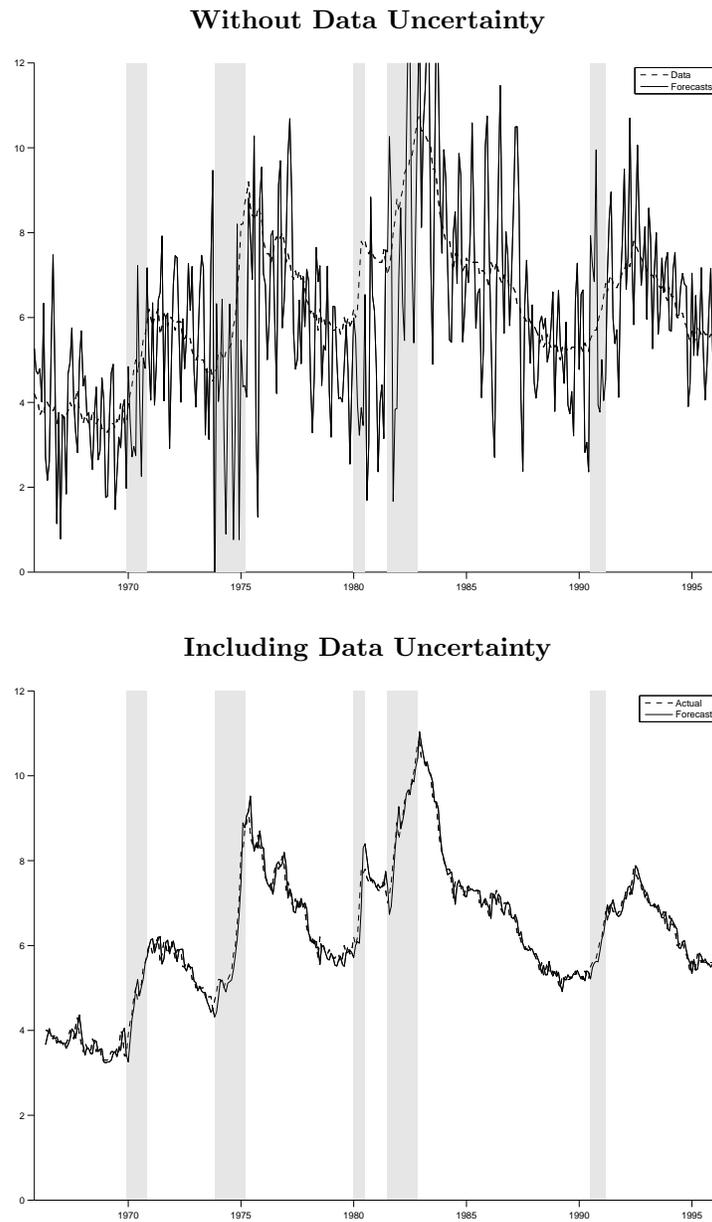


Including Data Uncertainty



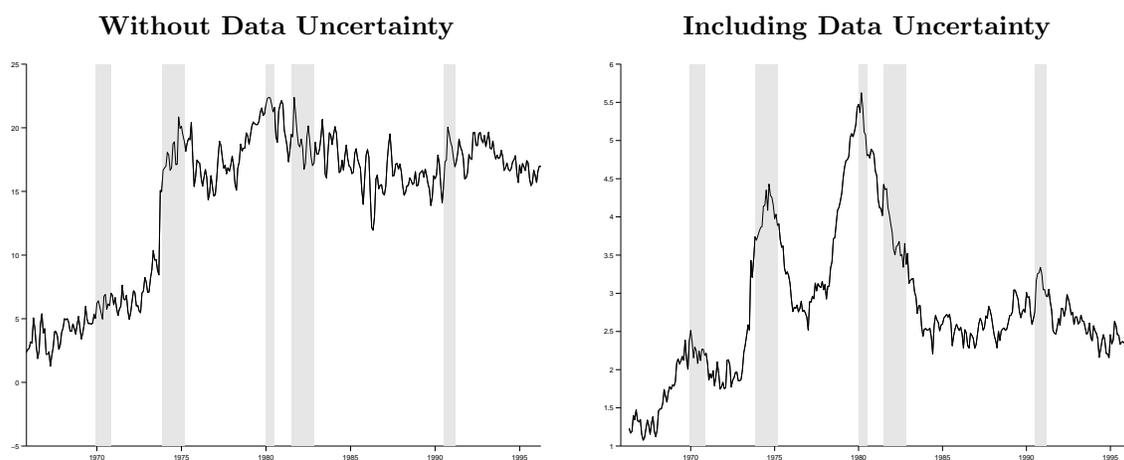
Notes: Real-time CPI inflation 1960–1996 as 12-month-ended growth rate. Model-predicted Federal Reserve optimal policy, from model without data uncertainty (top) and model including data uncertainty (bottom). NBER recessions shaded.

Figure 3: UNEMPLOYMENT RATE VS. FEDERAL RESERVE FORECASTS



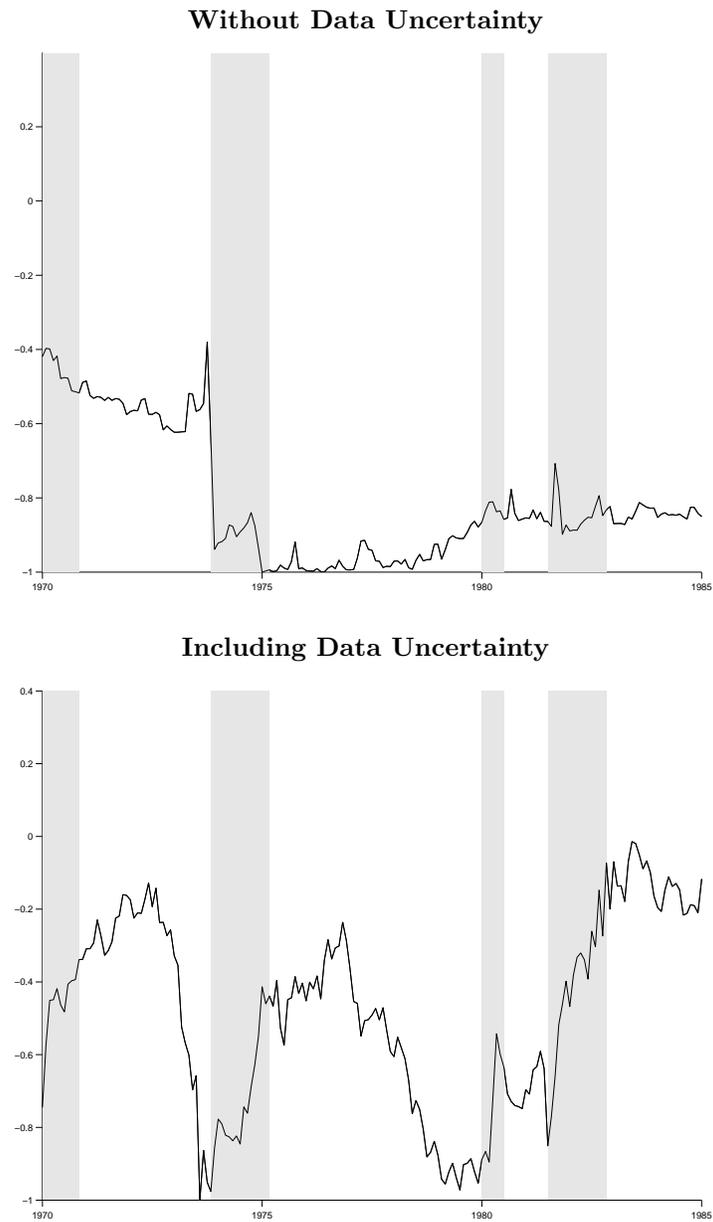
Notes: Real-time civilian unemployment rate 1960–1996, in percentage. Model-predicted step-ahead unemployment Federal Reserve unemployment rate forecasts. NBER recessions shaded.

Figure 4: PHILIPS CURVE CONSTANT



Notes: Estimated constant in the Philips curve. Note – vertical scales are not identical. NBER recessions shaded.

Figure 5: EVOLUTION OF THE PHILIPS CURVE TRADE-OFF



Notes: Estimated evolution of Philips curve trade-off, from sum of Philips curve parameters on inflation, set to normalized scale. NBER recessions are shaded.