Technical Appendices and Additional Results for "Inflation-Gap Persistence in the U.S.""

Timothy Cogley,* Giorgio E. Primiceri,† and Thomas J. Sargent‡

Revised: February 2009

Here we provide technical details about our estimation procedure and report additional results on post-WWII U.S. inflation-gap persistence. This supplement is not self-contained, so readers are advised to read the main paper first.

A Markov chain Monte Carlo algorithm for simulating the VAR posterior

For the VAR, the posterior density is[1]

$$p(\theta^T, H_y^T, H_s^T, B_y, B_s, \sigma_y, \sigma_s | Y^T). \tag{1}$$

The state and measurement innovation variances are defined as

$$\begin{aligned} Q_t &= (B_s^{-1})' H_{st} (B_s^{-1}), \\ R_t &= (B_y^{-1})' H_{yt} (B_y^{-1}), \end{aligned} \tag{2}$$

respectively, where H_{st} and H_{yt} are diagonal matrices with univariate stochastic volatilities along the main diagonal and B_s and B_y are triangular matrices with ones along the main diagonal and static covariance parameters below. The univariate stochastic volatilities are geometric random walks; the vectors σ_s and σ_y list their innovation variances. The notation x^T represents the complete history of x_t.

We use a 'Metropolis-within-Gibbs' algorithm to simulate the posterior. The parameters are partitioned into 7 blocks:

---

*New York University. Email: tim.cogley@nyu.edu.

†Northwestern University, CEPR, and NBER. Email: g-primiceri@northwestern.edu.

‡New York University and Hoover Institution, Stanford University. Email: ts43@nyu.edu.

[1]The MCMC algorithm for the univariate AR is a special case of that for the VAR.

- $\theta^T | H_y^T, H_s^T, B_y, B_s, \sigma_y, \sigma_s, Y^T$
- $H_y^T | \theta^T, H_s^T, B_y, B_s, \sigma_y, \sigma_s, Y^T$
- $B_y | \theta^T, H_y^T, H_s^T, B_s, \sigma_y, \sigma_s, Y^T$
- $\sigma_y | \theta^T, H_y^T, H_s^T, B_y, B_s, \sigma_s, Y^T$
- $H_s^T | \theta^T, H_y^T, B_y, B_s, \sigma_y, \sigma_s, Y^T$
- $B_s | \theta^T, H_y^T, H_s^T, B_y, \sigma_y, \sigma_s, Y^T$
- $\sigma_s | \theta^T, H_y^T, H_s^T, B_y, B_s, \sigma_y, Y^T$

After substituting Q_t for Q, the samplers for the first four blocks are identical to those in Cogley and Sargent (2005); details can be found there. Those for the last three blocks – which pertain to the state innovation variance Q_t – are isomorphic to the three blocks for the measurement innovation variance R_t. Thus, the appendices in Cogley and Sargent (2005) cover those blocks as well.

We executed 100,000 scans of the chain and diagnosed convergence by inspecting recursive mean plots of the parameters. We discarded the first 50,000 scans to allow for burn in. The results reported in the text are based on the remaining 50,000 scans.

A.1 Priors for the VAR

The priors are similar to those in Cogley and Sargent (2005). We assume that the hyperparameters and initial value of the drifting parameters are independent across blocks, so that the joint prior factors into a product of marginal priors. Each of the marginal priors is selected from a family of natural conjugate priors and is specified to proper yet weakly informative.

The unrestricted prior for the initial state is Gaussian,

$$f(\theta_0) \propto \mathcal{N}(\bar{\theta}, \bar{P}), \tag{3}$$

where $\bar{\theta}$ and $\bar{P}$ are the OLS point estimate and asymptotic variance, respectively, based on a training sample covering the period 1948-58. Because the training sample is short, the asymptotic variance is large, making the prior weakly informative for θ_0.

Priors for the blocks governing R_t are also calibrated to put considerable weight on sample information. The prior for H_{y0}^{ii} is log-normal,

$$f\left(\ln H_{y0}^{ii}\right) = \mathcal{N}(\ln R_0^{ii}, 10), \tag{4}$$

where $\ln R_0^{ii}$ is the estimate of the log of residual variance of variable i from the preliminary sample. A variance of 10 is huge on a log scale and allows a wide range of values for h_{i0}. As is the case for θ_0, the prior mean for H_{y0} is no more than a ballpark number surrounded by considerable uncertainty.

Similarly, the prior for β_y is normal with mean zero and a large variance,

$$f(\beta) = \mathcal{N}(0, 10000 \cdot I). \tag{5}$$

Lastly, the prior for σ_{yi}^2, the variance of the stochastic volatility innovations, is inverse-gamma

$$f(\sigma_i^2) = IG(\delta_i/2, 1/2), \tag{6}$$

with scale parameter $\delta = 0.0001$ and degree-of-freedom parameter equal to 1. This distribution is proper but has fat tails.

The priors for the blocks governing Q_t parallel those for R_t. The prior for H_{Q0}^i is also log-normal,

$$f\left(\ln H_{Q0}^{ii}\right) = \mathcal{N}(\ln Q_0^{ii}, 10), \tag{7}$$

where $Q_0 = \gamma^2 \bar{P}$ is calibrated in the same way as in Cogley and Sargent (2005). Similarly, the priors for β_Q and σ_Q have the same form as those for β_y and σ_y. We just alter the dimensions so that they conform to H_{Qt} instead of H_{Rt}. The prior mean for H_{Q0} induces only a slight degree of time variation in θ_t, but in other respects the priors are sufficiently uninformative that they permit a wide range of outcomes for Q_t.

B Markov chain Monte Carlo algorithm for simulating the DSGE posterior

As in An and Schorfheide (2006), we use a Metropolis-Hastings algorithm to simulate the posterior distribution of the coefficients of the DSGE model. Let y^T denote the set of available data and α the vector of coefficients of the DSGE model. Moreover, let $\alpha^{(j)}$ denote the j^{th} draw from the posterior of α. The subsequent draw is obtained by drawing a candidate value, $\tilde{\alpha}$, from a Gaussian proposal distribution with mean $\alpha^{(j)}$ and variance $s \cdot V$. We then set $\alpha^{(j+1)} = \tilde{\alpha}$ with probability equal to

$$\min\left\{1, \frac{p\left(\tilde{\alpha}|y^T\right)}{p\left(\alpha^{(i)}|y^T\right)}\right\}. \tag{8}$$

If the proposal is not accepted, we set $\alpha^{(j+1)} = \alpha^{(j)}$.

The posterior distribution for α, $p\left(\alpha|y^T\right)$, can be computed multiplying the prior density by the likelihood function. Because the DSGE model has a linear-Gaussian state-space representation, the likelihood function can be evaluated using the prediction-error decomposition and the Kalman filter.

The algorithm is initialized around the posterior mode, found using a standard maximization algorithm. We set V to the inverse Hessian of the posterior evaluated at the mode, while s is chosen in order to achieve an acceptance rate approximately equal to 25 percent. We run two chains of 70,000 draws and discard the first 20,000 to allow convergence to the ergodic distribution.

C A univariate autoregression with drifting parameters and stochastic volatility

Stock and Watson's (2007) univariate unobserved-components model makes inflation the sum of a driftless random walk and a martingale-difference error. Their model highlights the importance of drift in trend inflation, but it imposes that the inflation gap $g_t \equiv \pi_t - \tau_t$ is serially uncorrelated for all t. Because of this restriction, Stock and Watson's model is not a suitable vehicle for investigating whether inflation-gap persistence has changed over time.

Thus, as a modest extension, we add a lag of inflation to Stock and Watson's measurement equation and estimate a univariate $AR(1)$ model with drifting parameters and stochastic volatility. For this model, $y_t = \pi_t, X_t = [1, \pi_{t-1}]', \theta_t = [\mu_t, \rho_t]', B_y = 1$, and B_s is a $2x2$ matrix of the form of equation (6) in the main paper.

Notice that the parameter μ_t in the measurement equation is an intercept rather than a measure of trend inflation. Accordingly, we now approximate trend inflation by $\tau_t \approx \mu_t/(1 - \rho_t)$. To a first-order approximation, this is also a driftless random walk.

The parameter ρ_t governs the degree of inflation-gap persistence. We constrain ρ_t to be less than one in absolute value at all dates. Having assumed that trend inflation is a driftless random walk, the stability constraint on ρ_t just rules out a second unit or explosive root in inflation.

Figure 1 portrays the posterior median and interquartile range for ρ_t. for the two inflation measures. For GDP inflation, the inflation gap is moderately persistent throughout the sample. The median estimate for ρ_t was around 0.55 in the early 1960s. It increased gradually to 0.7 by 1980, and then fell in two steps in the early 1980s and early 1990s, eventually reaching a value of 0.3 at the end of the sample. These estimates imply half-lives of 3.5, 5.8, and 1.7 months, respectively. For PCE inflation, the gap was initially less persistent, with an autocorrelation of 0.3, but otherwise movements in ρ_t are similar to those for GDP inflation.

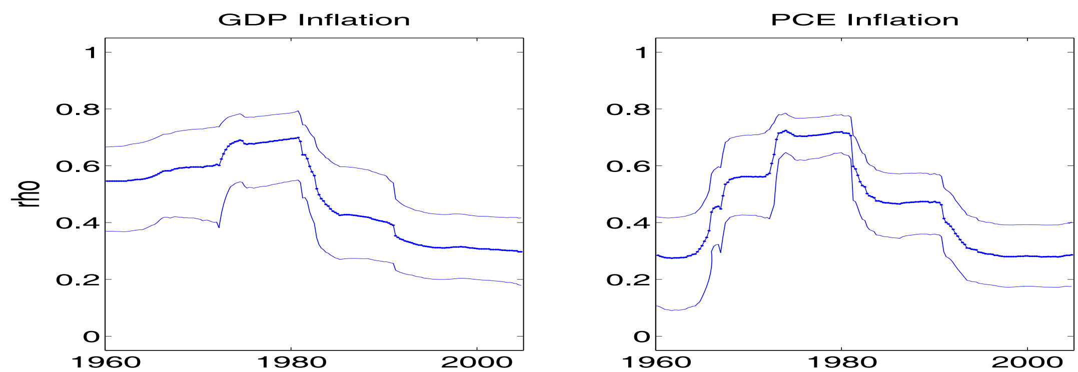

Figure 1: Posterior Median and Interquartile Range for ρ_t

To assess whether changes in ρ_t are statistically significant, we proceed as in the main text by inspecting the joint distributions for 1960.Q4-1980.Q4 and 1980.Q4-2004.Q4. Figure 2 portrays the results, with outcomes for GDP inflation shown in the left-hand column and those for PCE inflation in the right. The top row depicts the joint distribution for ρ_{1980} and ρ_{2004}, with values for 1980 plotted on the x-axis and those for 2004 on the y-axis. Combinations clustered near the 45 degree line represent pairs for which there was little or no change. Those below the 45 degree line represent a decrease in persistence ($\rho_{1980} > \rho_{2004}$), while those above represent increasing persistence. Similarly, the bottom row illustrates the joint distribution for ρ_{1960} and ρ_{1980}, with values for 1960 plotted on the x-axis and those for 1980 on the y-axis.

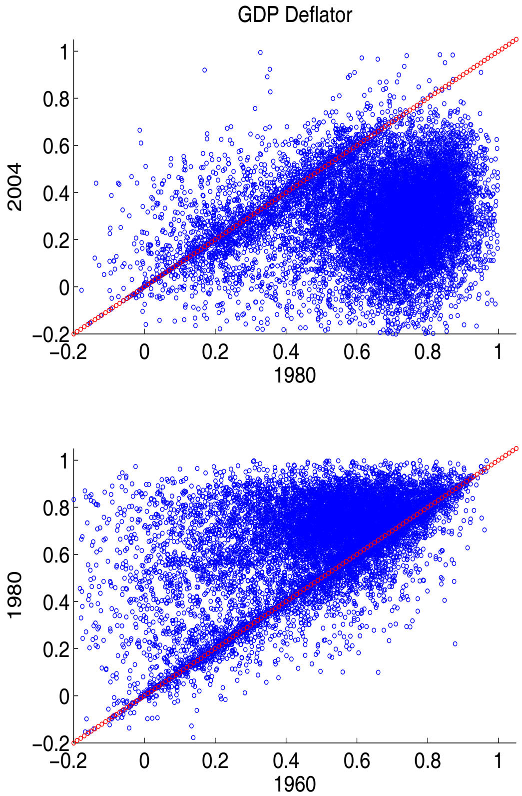

Figure 2: Joint Distributions for ρ_t, 1960-80 and 1980-2004

A number of alternative perspectives can be represented on these graphs. Stock and Watson assume $\rho_t = 0$, so the point $(0, 0)$ represents their model. There are some realizations in the neighborhood of the origin, but most of the probability mass lies elsewhere. The second column of table 1 reports the probability that ρ_t is close to zero in both periods, where 'close' is defined as $|\rho| < 0.1$. For GDP inflation, this comes out to 1.2 and 1.7 percent, respectively, for the two pairs of years. For PCE inflation, the probabilities are 0.006 and 0.008, respectively. This finding motivates our extension of their model.

Table 1: Posterior Probabilities

GDP Inflation

pair	Stock-Watson	$\|\Delta\rho\| < 0.05$	High, No Change	Changing ρ
1980, 2004	0.012	0.122	0.001	0.894
1960, 1980	0.017	0.384	0.027	0.758

PCE Inflation

pair	Stock-Watson	$\|\Delta\rho\| < 0.05$	High, No Change	Changing ρ
1980, 2004	0.006	0.056	0.001	0.959
1960, 1980	0.008	0.066	<0.001	0.956

To calculate the probability that inflation-gap persistence is approximately unchanged, we define a neighborhood along the 45 degree line and count the fraction of points that lie within. As figure 2 shows, the posterior distributions attach considerable probability mass to a ridge clustered tightly along the 45 degree line. How much probability is near that ridge depends on how a neighborhood is defined. For example, suppose we define 'little change' by the neighborhood $|\Delta\rho| < 0.05$. For GDP inflation, the 45-degree ridges are thickly populated, and the posterior probability that $|\Delta\rho| < 0.05$ comes to 12 and 38 percent, respectively, for the two pairs of years. For PCE inflation, the 45-degree ridges are more sparsely populated, and posterior probabilities are 5.6 and 6.6 percent, respectively. Obviously these probabilities would be higher if we widened the neighborhood and lower if we narrowed it, but the point is that the probability is nontrivial even for a narrowly defined interval along the 45 degree line. For the GDP deflator, the notion that univariate inflation-gap persistence is approximately constant cannot be rejected at the 10 percent level, while for PCE inflation that hypothesis can be rejected at the 10 percent level but not at the 5 percent level.

If we examine the ridges more closely, we notice that the scatterplots are densest along the ridge for low values of ρ and that they become sparse for high values. Thus, the notion that inflation-gap persistence is *both* unchanging *and* high has little support. For example, suppose we define 'high persistence' as a half-life of 1 year or more ($\rho \geq 0.8409$). For GDP inflation, the probability of high and unchanging persistence is less than one-tenth of 1 percent for 1980-2004 and 2.7 percent for 1960-1980. For PCE inflation, the probabilities are one-tenth of one percent or less. Inflation-gap persistence might have been high (especially during the Great Inflation), or it might have been unchanged, but it is unlikely that it was both. As noted above, the notion that persistence is both high and unchanging really applies to *inflation* – because of drift in τ_t – but not to the inflation *gap*.

In the top-left panel of figure 2, the largest probability mass of points – a bit less than 90 percent – lies below the 45 degree line. For combinations in this region, $\rho_{1980} > \rho_{2004}$, so this represents the probability of declining GDP inflation-gap persistence. We interpret this as substantial though not decisive evidence of a decline in persistence. Similarly, in the bottom-left panel, the preponderance of the combinations – approximately 75 percent – lie above the 45 degree line and are consistent with the idea that the inflation gap became more persistent between 1960 and 1980. Stronger evidence emerges for PCE inflation. The probability of an increase in ρ_t between 1960 and 1980 is 0.956, and the probability of a decline after 1980 is 0.959.

Thus, for GDP inflation the univariate evidence is mixed. The preponderance of the joint distribution points to a rise and then a decline in persistence, but there is enough mass along the 45 degree ridge to support the idea that inflation-gap persistence has not changed. For PCE inflation, there is significant evidence of a rise then a fall in inflation-gap persistence.

D Comparison with Atkeson-Ohanian findings

Stock and Watson interpret a result of Atkeson and Ohanian (2001) in terms of the changing time-series properties of inflation. Atkeson and Ohanian studied the predictive power of backward-looking Phillips-curve models during the Volcker-Greenspan era and found that Phillips-curve forecasts were inferior to a naive forecast that equates expected inflation over the next 12 months with the simple average of inflation over the previous year. Stock and Watson show that Phillips-curve models were more helpful during the Great Inflation, and they account for the change by pointing to two features of the data. First, like many macroeconomic variables, unemployment became less volatile after the mid-1980s. Hence there is less variation in the predictor. Second, the coefficients linking unemployment and other activity variables to future inflation have also declined in absolute value, further muting their predictive power.

Our VARs share these characteristics. In figure 3, we illustrate how news about unemployment alters forecasts of inflation. At each date, we imagine that forecasters start with information on inflation, unemployment, and the nominal interest rate through date $t-1$ and then see a one-sigma innovation in unemployment. They revise their inflation forecasts in light of the unemployment news. Because the VAR innovations are correlated, the forecast revision at horizon j is[2]

$$FR_{jt} = e_{\pi} A_t^j E(\varepsilon_{zt} | \varepsilon_{ut}) \sigma_{ut}. \tag{9}$$

Since the innovations are conditionally normal and the unemployment innovation is scaled to equal σ_{ut}, $E(\varepsilon_{zt}|\varepsilon_{ut}) = cov(\varepsilon_{zt}, \varepsilon_{ut})/\sigma_{ut}$. The figure portrays the median and interquartile range for forecast revisions at horizons of 1, 4, and 8 quarters.

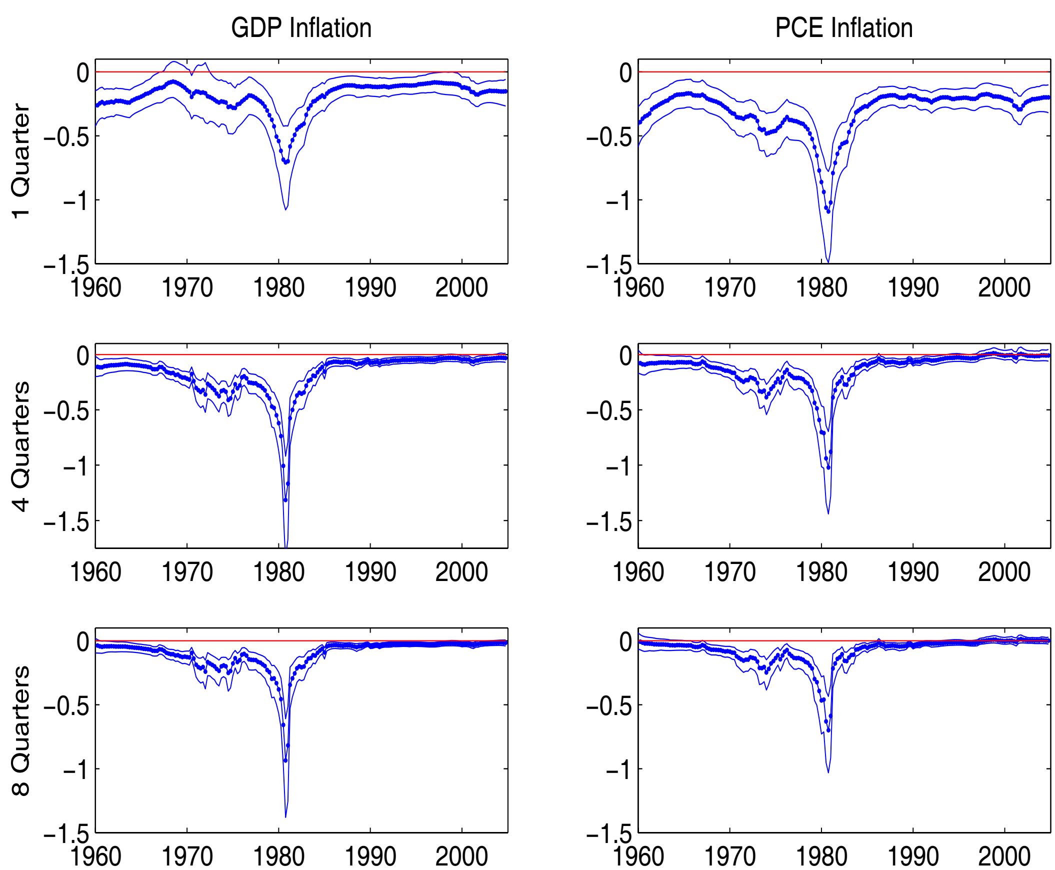

Figure 3: How Unemployment News Alters Expected Inflation

For the most part, a positive innovation in unemployment reduces expected inflation. Furthermore, in the 1970s and early 1980s, the magnitude of forecast revisions

[2]This follows from another anticipated-utility approximation.

was substantial. For instance, according to the median estimates, a one-sigma innovation in unemployment would have reduced expected inflation 4 quarters ahead by close to 50 basis points in the mid-1970s and by approximately 1 to 1.5 percentage points at the time of the Volcker disinflation. After the mid 1980s, however, the sensitivity of inflation forecasts to unemployment news was more muted. During the Greenspan era, a one-sigma innovation in unemployment would have had essentially no influence at all on inflation forecasts one or two years ahead.

As Stock and Watson point out, these outcomes reflect both that unemployment innovations are less volatile and that inflation forecasts are less sensitive to innovations of a given size. Figure 4 depicts the posterior median and interquartile range for σ_{ut}, the standard deviation of innovations to unemployment. The magnitude of unemployment innovations was largest at the beginning of the sample and around the time of the Volcker disinflation, but it declined sharply after the mid 1980s. One reason why unemployment news has become less relevant for inflation forecasting is that there is less of it.

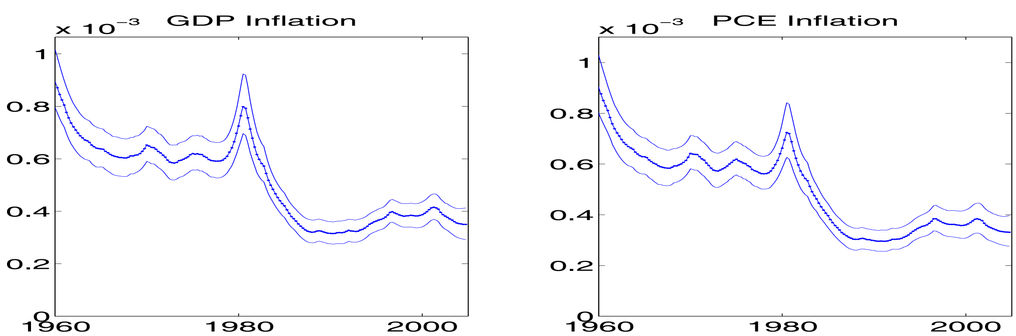

Figure 4: Standard Deviation of Unemployment Innovations

But this is not the whole story. Figure 5 adjusts for changes in the innovation variance by showing forecast revisions for the time-series average of the median estimate of σ_{ut} shown in figure 4. This holds the size of the hypothetical unemployment innovation constant across dates. Although less pronounced, the pattern shown here is similar to that depicted in figure 3 (the two figures are graphed on the same scale). Hence figure 3 cannot be explained solely by changes in σ_{ut}. Especially at horizons of a 4 or 8 quarters, inflation forecasts have also become less sensitive to a given amount of unemployment news than they were during the Great Inflation.

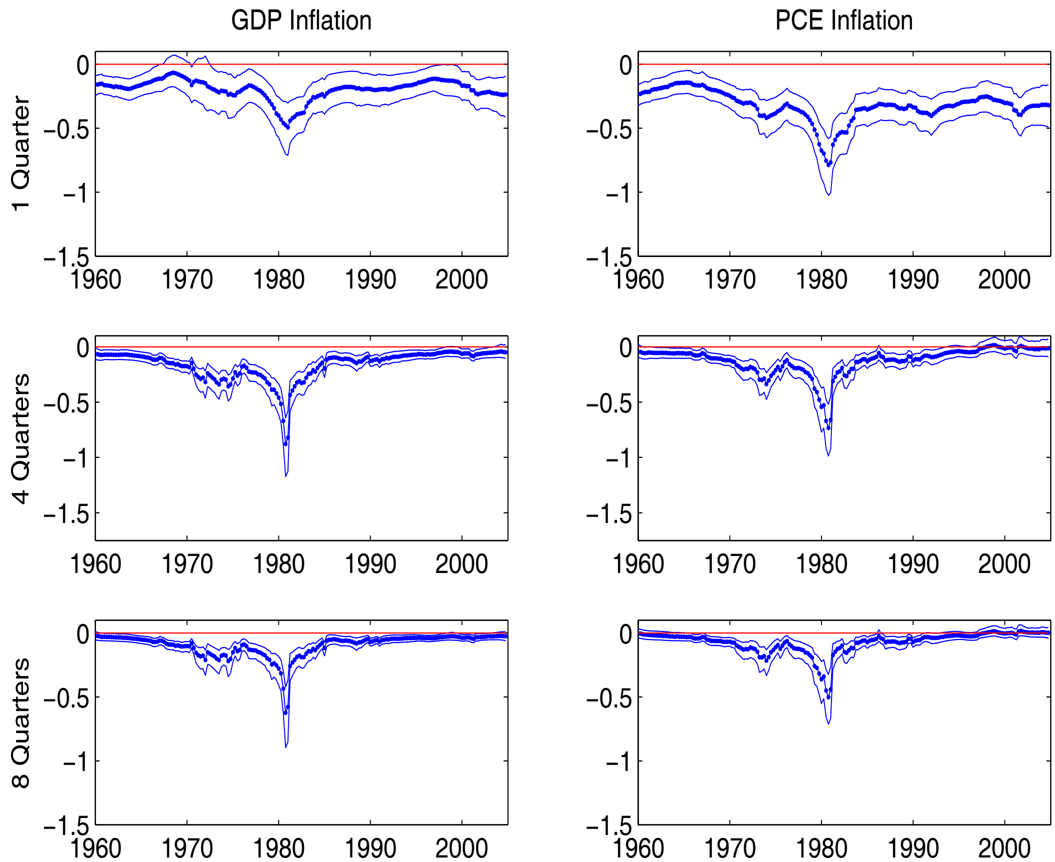

Figure 5: Forecast Revisions with σ_u Held Constant

Ironically, the decreased predictive power of unemployment innovations for inflation coincided with a return of the Phillips correlation. Figure 6 portrays a number of conditional and unconditional correlations for inflation and unemployment.[3] The unconditional correlation – shown in the bottom row – was negative prior to the 1970s, but it turned positive during the Great Inflation. A negative correlation reappeared after the Volcker disinflation and has hovered around -0.25 ever since.

The other rows of the figure depict conditional correlations at forecast horizons of 1, 4, and 8 quarters. The 1- and 4-quarter ahead forecasts are most relevant for reconciling conventional wisdom with Atkeson and Ohanian. At these horizons, conditional correlations have indeed been negative throughout the sample, peaking in

[3] These were also calculated using anticipated-utility approximations.

magnitude at the time of the Volcker disinflation. They are smaller now than in the past, but at the 4-quarter horizon the correlation is still around -0.25. Nevertheless, these conditional correlations are irrelevant for prediction because they summarize *unexpected* comovements in the two variables. That prediction errors in unemployment are inversely related with prediction errors in inflation tells us little about forecastable movements in the two variables. Thus, Atkeson and Ohanian's observations about predictability can coexist comfortably with conventional views about Phillips correlations.

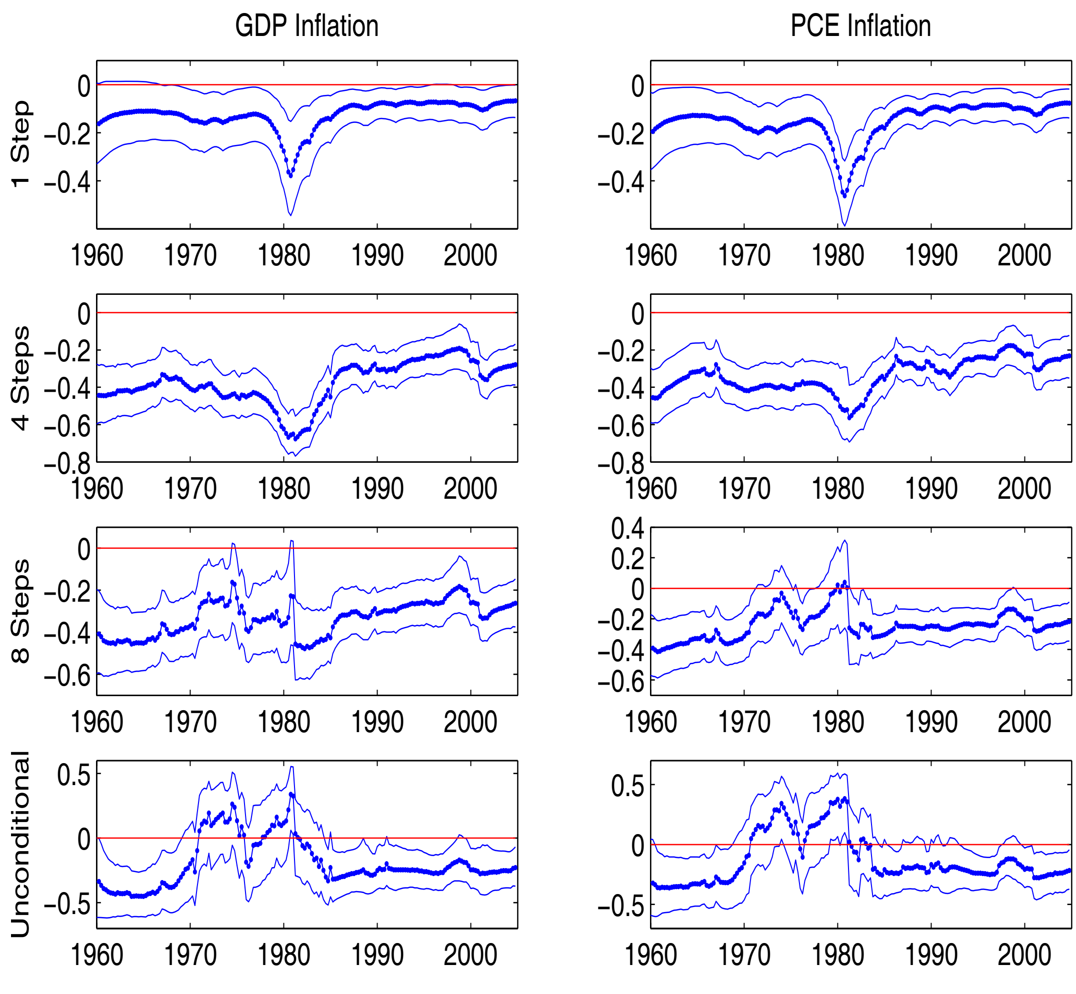

Figure 6: Conditional and Unconditional Phillips Correlations

Contrary to Atkeson and Ohanion, figure 3 suggests that some short-term predictability remains at the end of the sample. Two caveats should be kept in mind, however. One is that our calculations involve pseudo forecasts that depend on data and estimates through the end of the sample, while Atkeson and Ohanian look at real-time, out-of-sample forecasts. Presumably this matters only slightly at the end of the sample, but more for earlier periods.

The other caveat is that there is substantial uncertainty about R^2_{2004}. We can state with confidence that R^2_{2004} is smaller than R^2_{1980}, but that is mainly because the posterior for R^2_{1980} clusters tightly near 1. It is harder to say how predictable inflation is at the end of the sample. At the 1-quarter horizon, the probability that R^2_{2004} exceeds 0.25 is 0.904 for GDP inflation and 0.924 for PCE inflation. Thus, although our estimates suggest more predictability than those of Atkeson and Ohanian, the fact that the posteriors portrayed in figures 10 and 11 assign non-negligible probability to values of R^2_{2004} near zero provides some support for their point of view.

E A structural interpretation of the Atkeson-Ohanian findings

Finally, we examine what our structural model has to say about Atkeson and Ohanian's findings. Table 2 addresses the results of Atkeson and Ohanian (2001). Here we report the model-implied slope β of the Phillips curve,

$$E_t\left(\widehat{\overline{\pi}}_{4,t+4} - \widehat{\overline{\pi}}_{4,t}\right) = \beta\left(\hat{Y}_t - \hat{Y}^*_t\right).$$

We omit a constant because the model-generated variables in the regression all have mean zero. Except for the fact that we replace the unemployment rate with the output gap, this is the regression estimated by Atkeson and Ohanian (2001). Consistent with their results, those of Stock and Watson, and our own results reported above, our DSGE model implies a substantial decline in the predictive power of real-activity variables in a conventional Phillips curve regression after the Volcker disinflation. In our model, the coefficient on the output gap falls by 70 percent.

Table 2: Implications of the DSGE Model for a Phillips-Curve Regression

	Slope (β)
1960.Q1-1979.Q3	0.132
1982.Q1-2006.Q4	0.040
Percent Change	-70

Table 3 examines how counterfactual changes in monetary policy and private-sector parameters contribute to the flattening of the slope in an Atkeson-Ohanian regression. As in the main text, the numbers recorded here represent the proportion of the total change across subsamples accounted for by the hypothetical structural shift,

$$100 \times \frac{counterfactual\ change}{total\ change}.$$

Positive numbers signify that the counterfactual goes in the same direction as the total change, and negative numbers mean that it goes in the opposite direction.

Table 3: Counterfactual Exercises Based on the DSGE Model

Coefficients	Slope (β
Policy 2, Private 1	-94
σ_*	-46
ϕ_π	-26
Private 2, Policy 1	125
ρ_θ	121

The DSGE model predicts a total decline in β from 0.13 to 0.04 across the two subsamples. In this case, the relative importance of better policy and better luck are reversed. Changes in private sector parameters go in the right direction and overpredict the total decline. After a mark-up shock, the output gap and future changes in the inflation rate comove positively. The decline in the persistence and unconditional volatility of the mark-up shock after 1982 reduces this positive comovement and results in a lower estimate of the slope coefficient. But changes in policy parameters go in the wrong direction and predict a substantial increase in β. Thus, for a complete picture of the change in inflation outcomes, both private and policy factors are needed.

References

An, S. and F. Schorfheide, 2007, Bayesian Analysis of DSGE Models, *Econometric Reviews* 39, 1357-1373.

Atkeson, A. and L.E. Ohanian, 2001, Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review* 25, 2-11.

Cogley, T. and T.J. Sargent, 2005, Drifts and Volatilities: Monetary Policies and Outcomes in the Post War U.S., *Review of Economic Dynamics* 8, 262-302.

Stock, J.H. and M.W. Watson, 2007, Why Has Inflation Become Harder to Forecast?, *Journal of Money Credit and Banking* 39(1), 3-33.