NBER WORKING PAPER SERIES

WHY HAS U.S. INFLATION BECOME HARDER TO FORECAST?

James H. Stock
Mark W. Watson

Working Paper 12324
http://www.nber.org/papers/w12324

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2006

We thank Jonas Fisher for bringing several of the issues discussed in this paper to our attention in a 1999 conversation, Luca Benati for (more recent) helpful suggestions, and Matthew Shapiro, Robert Gordon, and two anonymous referees for helpful comments on an earlier draft. This research was funded in part by NSF grant SBR-0214131. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Why Has U.S. Inflation Become Harder to Forecast?
James H. Stock and Mark W. Watson
NBER Working Paper No. 12324
June 2006
JEL No. C53, E37

ABSTRACT

Forecasts of the rate of price inflation play a central role in the formulation of monetary policy, and forecasting inflation is a key job for economists at the Federal Reserve Board. This paper examines whether this job has become harder and, to the extent that it has, what changes in the inflation process have made it so. The main finding is that the univariate inflation process is well described by an unobserved component trend-cycle model with stochastic volatility or, equivalently, an integrated moving average process with time-varying parameters; this model explains a variety of recent univariate inflation forecasting puzzles. It appears currently to be difficult for multivariate forecasts to improve on forecasts made using this time-varying univariate model.

James H. Stock
Department of Economics
Harvard University
Littauer Center M4
Cambridge, MA 02138
and NBER
james_stock@harvard.edu

Mark W. Watson
Department of Economics
Princeton University
Princeton, NJ 08544-1013
and NBER
mwatson@princeton.edu

1. Introduction

The rate of price inflation in the United States has become both harder and easier to forecast, depending on one's point of view. On the one hand, inflation (along with many other macroeconomic time series) is much less volatile than it was in the 1970s or early 1980s, and the root mean squared error of naïve inflation forecasts has declined sharply since the mid-1980s. In this sense, inflation has become easier to forecast: the risk of inflation forecasts, as measured by mean squared forecast errors (MSFE), has fallen. On the other hand, the relative improvement of standard multivariate forecasting models, such as the backwards-looking Phillips curve, over a univariate benchmark has been less in percentage terms since the mid-1980s than before. This point was forcefully made by Atkeson and Ohanian (2001) (henceforth, AO), who found that, since 1984 in the U.S., backwards-looking Phillips curve forecasts have been inferior to a naïve forecast of average twelve-month inflation by its average rate over the previous twelve months. In this sense, inflation has become harder to forecast, at least, it has become much more difficult for an inflation forecaster to provide value added beyond a univariate model. One can speculate on economic reasons why this might be so, but a first step in informing such speculation is pinning down what, precisely, have been the changes in the inflation process that led to these changing properties of inflation forecasts.

This paper proposes a parsimonious model of the changes in the univariate process for postwar U.S. quarterly inflation, in which inflation is represented as the sum of two components, a permanent stochastic trend component and a serially uncorrelated transitory component. Since the mid 1950s, there have been large changes in the variance of the permanent disturbance, whereas the variance of the transitory disturbance has remained essentially constant. According to our estimates, the standard deviation of the permanent disturbance was moderate – for GDP inflation, approximately 0.5 percentage points at an annual rate – from the mid 1950s through approximately 1970; it was large, nearly 1.5 percentage points, during the 1970s through 1983; and it declined sharply in the mid 1980s to its value of the 1960s. Since 1990 it has declined further and now stands at a record low since 1954, less than 0.20 percentage points.

The time-varying trend-cycle model is equivalent to a time-varying first order integrated moving average (IMA(1,1)) model for inflation, in which the magnitude of the MA coefficient varies inversely with the ratio of the permanent to the transitory disturbance variance. Accordingly, the MA coefficient for inflation was small (approximately .25) during the 1970s but subsequently increased (to approximately .65 for the 1984-2004 period).

The time-varying trend-cycle model of the univariate inflation process succinctly explains the main features of the historical performance of univariate inflation forecasts. During the 1970s the inflation process was well approximated by a low order autoregression (AR), but in the mid 1980s the coefficients of that autoregressions changed and, even allowing for those changes, the low order autoregression became a less accurate approximation to the inflation process since 1984. The changing AR coefficients and the deterioration of the low-order AR approximation accounts for the relatively poor performance of recursive and rolling AR forecasts in the 1984-2004 sample. Moreover, it turns out that the AO year-upon-year forecast, represented as a linear combination of past inflation, is close to the optimal linear combination implied by the post-1984 IMA model at the four-quarter horizon, although this is not so at shorter horizons for the post-1984 period nor is it so at any horizon during the pre-1984 period, cases in which the AO forecasts perform relatively poorly.

This time-varying trend-cycle model also explains the excellent recent forecasting performance of an IMA model published by Nelson and Schwert (1977), which they estimated using data from 1953 to 1971. During the 1970s and early 1980s, the variance of the permanent component was an order of magnitude larger than it was in the 1950s and 1960s, and the Nelson-Schwert (1977) model did not provide good forecasts during the late 1970s and early 1980s. During the late 1980s and 1990s, however, the size of the permanent component fell back to its earlier levels, and the Nelson-Schwert (1977) model was again a good approximation.

The time-varying trend-cycle model provides a strategy for real-time univariate forecasting. Currently the Nelson-Schwert (1977) forecast is performing very well, and the AO forecast is performing nearly as well, at least at long horizons. But the inflation process has changed in the past, it could change again, and if it does, the performance of fixed-parameter models like AO and Nelson-Schwert will deteriorate. We therefore consider it imprudent to adopt a fixed-parameter inflation forecasting model either as a benchmark or for real-time forecasting. Instead, our pseudo-out-of-sample forecasting results suggest two approaches to time-varying trend-cycle models which could be effective in the face of such changes: an unobserved components model with stochastic volatility, implemented using a nonGaussian filter, and an IMA(1,1) model with moving average coefficient estimated using a ten-year rolling window of past observations. The rolling IMA(1,1) model is simpler, but adapts to changing parameters less quickly than, the unobserved components/stochastic volatility model.

The changing univariate inflation dynamics also help to explain the dramatic breakdown of recursive and rolling autoregressive distributed lag (ADL) inflation forecasts based on an activity measure. One reason for the deterioration in the relative performance of the ADL activity-based forecasts is that the variance of the activity measures has decreased since the mid-1980s (this is the "Great Moderation"), so in a sum-of-squares sense their predictive content, assuming no changes in coefficients, has declined. But the coefficients of the ADL models have also changed. Because the ADL forecasts generalize a univariate autoregression, they inherit the defects of the univariate AR forecasts in the second period. The evidence on the stability and statistical significance of the coefficients on lagged activity variables in the ADL is mixed, and sampling variability impedes making sharp statements about the stability of the Phillips curve after allowing for changes in the coefficients on lagged inflation. Although a complete analysis of Phillips curve forecasts that incorporate these time-varying coefficients is beyond the scope of this paper, we illustrate some implications of our univariate findings for multivariate analysis.

The literature on inflation forecasting and the empirical Phillips curve is too large to survey here comprehensively, but several recent papers which are closely related to this one are noteworthy. Fischer, Liu, and Zhou (2002) and Orphanides and Van Norden (2005) confirmed Atkeson and Ohanian's (2001) basic point that, since the mid-1980s, it has been quite difficult for inflation forecasts to improve on simple univariate models. Roberts (2004) identified a flattening of the Phillips curve and a change in the coefficients on lagged inflation in unemployment-rate Phillips curves occurring around 1984, the break date we focus on in much of our analysis. Clark and McCracken (2005) stress that the sampling variability of pseudo out-of-sample forecast comparison statistics is so large that the statistical case for the breakdown in Phillips curve forecasts is not watertight, despite their poor recent performance in economic terms. We too find considerable sampling variability in forecast comparison measures and return to this point below. Dossche and Everaert (2005), Harvey, Trimbur, and van Dijk (2005), and Leigh (2005) also implement unobserved components models of inflation (for different purposes), although their models omit the stochastic volatility that is the central part of our story.

The rest of the paper is organized as follows. Section 2 lays out the main forecasting facts and puzzles. Sections 3 – 5 examine changes in the univariate inflation process. Section 6 lays out some implications of the univariate results for activity-based Phillips curve forecasts. Section 7 concludes.

2. U.S. Inflation Forecasts: Facts and Puzzles

This section summarizes the performance of models for forecasting U.S. inflation using a pseudo out-of-sample forecast comparison methodology, with a focus on answering the question of whether inflation has become harder to forecast. One purpose of this section is to provide a consistent and concise summary of miscellaneous related results that appear elsewhere in the literature on inflation forecasting and volatility (see Ang, Bekaert, and Wei (2005), Atkeson and Ohanian (2001), Clark and McCracken

(2005), Stock and Watson (2002), and Tulip (2005); for complementary results for the UK, see Benati and Mumtaz (2005)). The section begins with a description of the data and the forecasting models, then turns to the results. To keep things simple, in this section we focus on split-sample results, comparing the period 1970:I – 1983:IV to the later period 1984:I – 2004:IV. The sample split date of 1984 coincides with estimates of the onset of the great moderation and is the split date chosen by Atkeson and Ohanian (2001). These split sample results convey the main facts about the changing behavior of inflation forecasts. In subsequent sections, we examine formal evidence for a break at an unknown date and consider methods that allow for continual rather than discrete changes in the inflation process and the forecasting relations.

2.1 Data

The paper focuses on GDP price index inflation. As a sensitivity analysis, results are also presented for the personal consumption expenditure deflator for core items (PCE-core), the personal consumption expenditure deflator for all items (PCE-all), and the consumer price index (CPI, the official CPI-U). We consider a number of activity variables: the unemployment rate (all, 16+, seasonally adjusted) (u), log real GDP (y), the capacity utilization rate, building permits, and the Chicago Fed National Activity Index (CFNAI) (for discussion of the choice of activity predictors see for example Stock and Watson (1999)). For series with revisions, the vintage as of May 2005 was used.

All empirical work uses quarterly data. Quarterly values for monthly series were computed by averaging the monthly values for the three months in the quarter; if logarithms are taken, they are logarithms of the average value of the monthly indexes.[1] For the main results, the full sample is from 1960:I through 2004:IV, with earlier data used to initialize regressions with lags; results that use a different sample are noted explicitly.

Some predictors appear in "gap" form, denoted (for example) as $ugap^{1\text{-}sided}$ and $ugap^{2\text{-}sided}$ for the one- and two-sided unemployment gaps. Gaps are computed as deviation of the univariate activity series (e.g. u) from a lowpass filter with pass band corresponding to periodicities of 60 quarters and higher. Two-sided gaps are computed as deviations from the symmetric two-sided MA(80) approximation to the optimal lowpass filter after padding the endpoints of the series with backcasts and forecasts computed from an estimated AR(4) model. One-sided gaps are computed using the same MA(80) filter replacing future observations with recursively constructed AR(4) forecasts. Two-sided gaps are useful for analyzing historical relationships but are not feasible for forecasting.

2.2 Forecasting Models and Pseudo Out-of-Sample Methodology

We begin by considering two univariate forecasting models and one multivariate forecasting model, implemented using different predictors. Let $\pi_t = 400\ln(P_t/P_{t-1})$, where P_t is the quarterly price index, and let h-period average inflation (at an annual rate) be $\pi_t^h = h^{-1}\sum_{i=0}^{h-1}\pi_{t-i}$. Adopt the notation that subscript $|t$ on a variable denotes the forecast

[1] The analysis was also performed using end-of-quarter aggregation with no important changes in the qualitative conclusions. Many of the coefficient values reported below are sensitive to the method of temporal aggregation (as they should be) but the magnitude and timing of the changes in parameters and the consequent conclusions about forecasting are not.

made using data through time *t*, for example $\pi^h_{t+h|t}$ is the forecast of π^h_{t+h} made using data through *t*.

AR(AIC). Forecasts are made using a univariate autoregression, specified in terms of the change of inflation with *r* lags, where *r* is estimated using the Akaike Information Criterion (AIC). Multistep forecasts are computed by the direct method, that is, by projecting *h*-period ahead inflation on *r* lags. Specifically, the *h*-step ahead AR(AIC) forecast was computed using the model,

$$\pi^h_{t+h} - \pi_t = \mu^h + \alpha^h(\mathrm{B})\Delta\pi_t + u^h_t, \tag{1}$$

where μ^h is a constant, $\alpha^h(\mathrm{B})$ is a lag polynomial written in terms of the backshift operator B, u^h_t is the *h*-step ahead error term, and the superscript *h* denotes the quantity for the *h*-step ahead regression. Note that this specification imposes that π_t has a unit root.

AO. Atkeson-Ohanian (2001) (AO) forecasted the average four-quarter rate of inflation as the average rate of inflation over the previous four quarters. They did not forecast at other horizons so there is some ambiguity in specifying the AO forecast at other horizons. Because the AO forecast is essentially a random walk forecast, and a random walk forecast is the same at all horizons, we extend the AO forecast to other horizons without modification. Thus the AO forecast is,

$$\pi^h_{t+h|t} = \pi^4_t = \frac{1}{4}(\pi_t + \ldots + \pi_{t-3})\,. \tag{2}$$

Backwards-looking Phillips curve (PC). The PC forecasts are computed as direct ADL forecasts, that is, by adding a predictor to (1) to form the autoregressive distributed lag (ADL) specification,

$$\pi_{t+h}^{h} - \pi_t = \mu^h + \alpha^h(\mathrm{B})\Delta\pi_t + \beta^h xgap_t + \delta^h(\mathrm{B})\Delta x_t + u_t^h, \tag{3}$$

where $xgap_t$ is the gap variable based on the variable x_t and Δx_t is the first difference of x_t. The lag lengths of $\alpha^h(\mathrm{B})$ and $\delta^h(\mathrm{B})$ are chosen by AIC. The PC forecast using u_t as the gap variable (so $u_t = xgap_t = x_t$) and Δu_t as Δx_t is denoted *PC-u*; this is the conventional backwards-looking Phillips curve specified in terms of the level of the unemployment rate with a constant NAIRU, omitting supply shock control variables. The forecasts *PC-Δu*, *PC-Δy*, *PC-ΔCapUtil*, *PC-ΔPermits*, and *PC-CFNAI* omit gap variables and only include the stationary predictors Δu, Δy, Δ capacity utilization, Δ building permits, and the CFNAI, respectively. The *PC-Δy* forecast, which uses only the growth rate of GDP as a predictor and omits a gap term, is the activity-based inflation forecast recommended by Orphanides and van Norden (2005).

Pseudo out-of-sample forecast methodology. All forecasts were computed using the pseudo out-of-sample forecast methodology, that is, for a forecast made at date t, all estimation, lag length selection, etc. was performed using only data available through date t. The forecasts in this section are recursive, so that forecasts at date t are based on all the data (beginning in 1960:I) through date t. The period 1960-1970 was used for initial parameter estimation. The forecast period 1970:I – 2004:IV was split into the two periods 1970:I – 1983:IV and 1984:I – 2004:IV.

2.3 Results

The results of the pseudo out-of-sample forecast experiment are summarized in Table 1, where the different panel of the table report results for the three inflation series. The first row in each panel reports the root mean square forecast error (RMSFE) of the benchmark AR(AIC) forecast in percentage points at an annual rate, at the indicated forecast horizon h. The remaining rows report the MSFE of the row forecast, relative to the AR(AIC) (so the relative MSFE of the AR(AIC) forecast is 1.00); an entry less than one indicates that the candidate forecast has a lower MSFE than the AR(AIC) benchmark.

Table 1 does not report standard errors for the relative MSFEs, however standard errors are reported for the univariate model MSFEs are reported in Table 4 in the next section. As is emphasized by Clark and McCracken (2005), the standard errors can be large, ranging from 0.05 to 0.20 for four-quarter ahead forecasts. Despite these large standard errors, the results in Table 1 suggest four conclusions.

Table 1
Pseudo Out-of-Sample Forecasting Results for GDP Inflation

Multivariate forecasting model: $\pi_{t+h}^h - \pi_t = \mu^h + \alpha^h(B)\Delta\pi_t + \beta^h xgap_t + \delta^h(B)\Delta x_t + u_t^h$

	1970:I – 1983:IV				1984:I – 2004:IV				$\frac{RMSFE_{84-04}^{h=4}}{RMSFE_{70-83}^{h=4}}$
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	
AR(AIC) RMSFE	1.72	1.75	1.89	2.38	0.78	0.68	0.62	0.73	
Relative MSFEs									
AR(AIC)	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	1.00	0.33
AO	1.95	1.57	1.06	1.00	1.22	1.10	**0.89**	**0.84**	0.30
PC-*u*	**0.85**	0.92	0.88	0.61	**0.95**	1.11	1.48	1.78	0.42
PC-Δ*u*	0.87	**0.87**	0.86	0.64	1.06	1.27	1.83	2.21	0.48
PC-$ugap^{1\text{-}sided}$	0.88	0.99	0.98	0.87	1.06	1.29	1.84	2.39	0.45
PC-Δ*y*	0.99	1.06	0.93	0.58	1.05	1.06	1.23	1.53	0.37
PC-$ygap^{1\text{-}sided}$	0.94	0.97	0.99	0.78	0.97	0.97	1.25	1.55	0.37
PC-CapUtil	0.85	0.88	**0.79**	**0.55**	0.95	1.01	1.35	1.52	0.43
PC-ΔCapUtil	1.02	1.00	0.87	0.64	1.03	1.10	1.30	1.51	0.40
PC-Permits	0.93	1.02	0.98	0.78	1.08	1.23	1.31	1.52	0.38
PC-ΔPermits	1.02	1.04	0.99	0.86	1.00	1.00	1.00	1.02	0.33
PC-CFNAI	.	.	.	.	1.11	1.27	1.86	2.25	.

Notes to Table 1: The first row of entries are root mean squared forecast errors (RMSFEs) of the AR(AIC) benchmark forecast. For the remaining rows, the first eight numerical columns report the MSFE of the forecasting model, relative to the AR(AIC) benchmark (hence AR(AIC) = 1.00). The multivariate forecasts are denoted PC-*x*, where *x* is the activity variable used in the autoregressive distributed lag model stated in the table header. For the *PC-u*, *PC-* $ugap^{1\text{-}sided}$, *PC-y*, *PC-*$ygap^{1\text{-}sided}$, *PC-CapUtil*, and *PC-Permits* forecasts, the variable $xgap_t$ is, respectively, u_t, $ugap_t^{1\text{-}sided}$, etc. For the remaining forecasts, *xgap* is omitted and Δx_t is given in the forecast name, e.g. for *PC-Δu*, $\Delta x_t = \Delta u_t$. In the PC forecasts, the lag lengths for $\alpha^h(B)$ and $\delta^h(B)$ were chosen independently by AIC, with between 0 and 4 lags, and the forecasts are direct (not iterated). The final column reports the reduction in RMSFE from the 1970-1983 period to the 1984-2004 period for the row forecasting method, at the four-quarter horizon. Bold entries denote the lowest MSFE for that period/horizon. All forecasts are pseudo out-of-sample. Results for the CFNAI are only computed for the second sample because of a shorter span of data availability.

1. *The RMSFE of forecasts of GDP inflation has declined. In this sense, inflation has become easier to forecast.* The magnitude of this reduction is striking. Whatever its other merits or demerits, the AR(AIC) forecast is simple to produce and has been a staple of economic forecasters for decades, and a forecaster using this method consistently from 1984 to 2004 would have RMSFEs of only 0.62 percentage points for annual GDP inflation, down from 1.89 percentage points over the 1970-1983 period, a reduction of two-thirds. As the final column of Table 1 demonstrates, even those forecasting models with performance that deteriorates in the second period, relative to the first, still exhibit reductions in forecast uncertainty of at least one-half.

2. *The relative performance of the Phillips curve forecasts deteriorated substantially from the first period to the second.* For example, during the 1970-1983 period at the four-quarter horizon, the PC-u forecast of GDP inflation outperformed the AR(AIC) benchmark (relative MSFE = .88), but during the 1984-2004 period it performed worse than the AR(AIC) benchmark (relative MSFE = 1.48). The change in relative performance is even larger at $h = 8$, but there are fewer nonoverlapping observations at this horizon so the $h = 8$ relative MSFEs have considerable sampling uncertainty (cf. Clark and McCracken (2005)). This deterioration of Phillips curve forecasts is found for all the activity predictors examined in the table. For example, at the eight quarter horizon, forecasts based on the capacity utilization rate had a RMSE

45% less than the AR(AIC) in the 1970-1983 sample but had a RMSE 52% greater in the 1984-2004 sample.

3. *The poor performance of the PC forecasts is not simply a consequence of failing to allow for a time-varying NAIRU or time-varying potential GDP*. The PC-Δu and PC- $ugap^{1\text{-}sided}$ specifications allow for a slowly time-varying NAIRU (in the case of PC-Δu, by omitting the level of u). In some cases these outperform the PC-u specification, but in other cases, especially in the post-1984 sample, they do worse. Whether one allows for a time-varying NAIRU or not, the PC forecasts are not competitive with either the AR(AIC) or AO forecasts in the 1984-2004 sample.

4. *The AO forecast substantially improves upon the AR(AIC) and Phillips curve forecasts at the four- and eight-quarter horizons in the 1984-2004 period, but not at shorter horizons and not in the first period*. The shift in relative performance is dramatic. For example, at the $h = 4$ horizon (the only horizon reported by Atkeson and Ohanian (2001)), in the first period the MSFE of the AO forecast, relative to the PC-u forecast, is 1.06/0.88 = 1.21, whereas in the second period this relative MSFE is .89/1.48 = 0.60.

A different way to summarize Table 1 is that inflation has become both easier and harder to forecast. On the one hand, inflation is easier to forecast in the sense that all these forecasting models have RMSFEs that are much smaller after 1984 than before. On the other hand, after 1984 it has been harder to be an inflation forecaster, in the sense that it is more difficult to improve upon simple univariate models, at least using activity-based backward-looking Phillips curves.

Evidently, there have been major changes in the univariate inflation process and in the bivariate process of inflation and its activity-based predictors, however these results do not indicate what those changes were or when they occurred.

3. Changes in the Univariate Inflation Process

The remainder of this paper explores a question raised by Table 1: what, specifically, have been the changes in the inflation process that led to the apparent changes in the relative and absolute performance of inflation forecasting models? In this section, we begin to answer this question by focusing on the univariate inflation process. This section continues to consider split-sample results with a 1984 break date; timing of changes is considered in Section 4.

3.1 Split-Sample Summary Statistics

There have been substantial changes in the autocorrelations and spectra of inflation. Some measures of persistence of the inflation process have changed, while others have not.

Volatility and autocorrelations. Table 2 presents the standard deviation and first eight autocorrelations of $\Delta\pi_t$ in both periods for GDP inflation. In the first period the only autocorrelation that is nonzero at the 10% level is the first (the t-statistic is 1.89); in

Table 2: Summary Statistics, GDP Inflation

	1960:I − 1983:IV	**1984:I − 2004:IV**
Standard deviation of $\Delta\pi_t$	1.30	0.91
Autocorrelation of $\Delta\pi_t$ at lag:		
1	-0.187 (0.102)	**-0.416** (0.109)
2	-0.148 (0.106)	-0.084 (0.127)
3	-0.006 (0.108)	-0.117 (0.127)
4	0.150 (0.108)	**0.395** (0.129)
5	-0.048 (0.110)	-0.268 (0.142)
6	-0.011 (0.110)	-0.020 (0.148)
7	-0.062 (0.110)	-0.000 (0.148)
8	0.001 (0.110)	0.304 (0.148)
Largest AR root of π_t	0.884 - 1.030	0.852 - 1.032

Notes to Table 2: The first row reports the standard deviation of the quarterly change of inflation $\Delta\pi_t$ (at an annual rate). The autocorrelations are for the indicated sample period, with standard errors in parentheses. Bold entries are significant at the 5% (two-sided) significance level. The final row reports the 90% confidence interval for the largest autoregressive root of inflation was computed using Stock's (1991) method of inverting the augmented Dickey-Fuller (ADF) test statistic (constant, no time trend).

the second period, the first autocorrelation is statistically significant at the 5% level, along with the fourth, perhaps reflecting some seasonality. In both periods, $\Delta\pi_t$ is negatively serially correlated (except for the positive fourth autocorrelation), with the first autocorrelation much larger in absolute magnitude (more negative) in the second period than the first.

Persistence. One measure of persistence is the magnitude of the largest autoregressive root of the levels process, in this case, the largest autoregressive root of inflation. As shown in Table 2, by this measure persistence did not change substantially between the two periods, either in a qualitative or quantitative sense. The confidence intervals are remarkably stable across series and time periods, being approximately (.85, 1.03); both confidence intervals include a unit root. By this measure, the persistence of inflation has not changed between the two periods. These split-sample confidence intervals are consistent with the results of the thorough analysis in Pivetta and Reis (2004), who report rolling and recursive estimates of the largest autoregressive root (and the sum of the AR coefficients) that are large and stably near one.

Although the largest AR root appears to be stably large, inflation persistence can nevertheless be viewed as having fallen: the fraction of the variance of $\Delta\pi_t$ explained by the persistent shocks and the fraction of the mass of the spectrum of $\Delta\pi_t$ near frequency zero are both considerably greater in the first period than in the second (see Appendix figure A.1).

3.2 The IMA/Trend-Cycle Model

The IMA(1,1) and unobserved components models. The apparent unit root in π_t and the negative first order autocorrelations, and generally small higher order autocorrelations, of $\Delta\pi_t$ suggest that the inflation process might be well described by the IMA(1,1) process,

$$\Delta\pi_t = (1 - \theta \mathrm{B})a_t, \tag{4}$$

where a_t is serially uncorrelated with mean zero and variance σ_a^2.

The IMA(1,1) model is equivalent to an unobserved components (UC) model in which π_t has a stochastic trend τ_t and a serially uncorrelated disturbance η_t:

$$\pi_t = \tau_t + \eta_t, \quad \eta_t \text{ serially uncorrelated } (0, \sigma_\eta^2) \tag{5}$$

$$\tau_t = \tau_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ serially uncorrelated } (0, \sigma_\varepsilon^2), \tag{6}$$

where $\mathrm{cov}(\eta_t, \varepsilon_t) = 0$.

Results. Table 3 presents estimates of the IMA(1,1) parameters, the implied UC parameters, and statistics testing the IMA(1,1) model against more general ARIMA models, for the pre- and post-1984 periods. Consistent with the changes in the autocorrelations, the MA parameter is considerably larger in the second period than in the first; consistent with the decline in the variance of inflation, the MA innovation has a smaller variance in the second period than in the first. This is true for all series.

Table 3. IMA(1,1) Model of $\Delta\pi_t$ and its Unobserved Components Representation, GDP Price Index

	1960:I − 1983:IV	1984:I − 2004:IV
(a) IMA parameters: $\Delta\pi_t = (1 - \theta B)a_t$		
θ	0.275 (.085)	0.656 (.088)
σ_a	1.261 (.070)	0.753 (.070)
(b) UC parameters		
σ_ε	0.914 (.118)	0.259 (.072)
σ_η	0.662 (.110)	0.610 (.068)
(c) p-values of Wald Tests of IMA(1,1) versus:		
ARIMA(1,1,1)	0.42	0.98
IMA(1,4)	0.40	0.13
(d) ARIMA(1,0,1) parameters: $(1 - \phi B)\pi_t = (1 - \theta B)a_t$		
ϕ	0.987 (.018)	0.989 (.011)
θ	−0.261 (.102)	−0.673 (.084)
(e) Tests for parameter stability		
t-statistic for $\sigma_{\varepsilon,70\text{-}83} = \sigma_{\varepsilon,84\text{-}04}$ (*p*-value)	–	–4.75 (<.001)
t-statistic for $\sigma_{\eta,70\text{-}83} = \sigma_{\eta,84\text{-}04}$ (*p*-value)	–	–0.41 (.684)
QLR: UC model (*p*-value)	–	31.99 (<.01)
QLR: AR(4) model (*p*-value)	–	4.23 (0.02)
(f) Variance decomposition of four-quarter inflation forecasts from the UC model		
4-quarter MSE	1.99	0.35
MSE due to:		
filtering error	0.32	0.13
trend shocks	1.57	0.13
transitory shocks	0.11	0.09

Notes to Table 3: Block (a) reports estimated parameters of the IMA(1,1) model (standard errors in parentheses), block (b) reports the corresponding parameters of the unobserved components model, block (c) reports tests of the IMA(1,1) specification against ARIMA models with more parameters, and block (d) reports estimates of ARIMA(1,0,1) models that do not impose a unit root in inflation; in all these blocks, standard errors are in parentheses. Block (e) reports tests for parameter stability, first one parameter at a time in the UC model with an imposed break in 1984 (shown as *t*-statistics), then the Quandt (maximal) likelihood ratio (QLR) statistic (df = 2) over all break dates in the inner 70% of the sample (the QLR statistic) for the UC model, and finally the (heteroskedasticity-robust) *F* – statistic version of the QLR statistic for an AR(4) model for $\Delta\pi$ (df=5). The *p*-values in parentheses in the final block take the 1984 break date as exogenous for the first two rows, but the QLR critical values allow for an endogenous break (Andrews (1993)). Block (f) reports a decomposition of the total four-quarter ahead forecast error variance (first row), based on the UC model, into the three components of filtering (signal extraction) error, future permanent disturbances, and future transitory disturbances.

Blocks (c)-(f) of Table 3 report various statistics assessing the fit and stability of the IMA(1,1) model. Wald tests of the null that the process is an IMA(1,1), against the alternative that it is a higher order process, fail to reject in both cases.

Although the unit root confidence intervals in Table 2 include one, the confidence intervals include values of the largest AR root that are relatively small, less than 0.9, so one might ask for additional evidence on the magnitude of the AR root in the context of a model with a moving average term. To this end, Table 3 reports ARIMA(1,0,1) models that do not impose a unit root in inflation. The point estimates are strikingly close to one, the smallest being .987. Because the distribution theory for the estimator of α is nonstandard when its true value is close to one, we rely on the confidence intervals in table 2 for formal inference about this root.

Block (e) of Table 3 reports tests of the hypothesis of parameter stability in the UC model. The hypothesis that the permanent innovation has the same variance in the two periods is strongly rejected, but the hypothesis that the transitory variance is the same is not. These tests are Chow tests which treat the 1984 break date as exogenously specified, which is inappropriate in that a large body of evidence about changes in the U.S. macroeconomy informed our choice of a 1984 break. To address this concern, the final two lines report the Quandt likelihood ratio (QLR) statistic, which is the maximum likelihood ratio test for a break over all possible break dates in the inner 70% of the full sample (1959 – 2004), first for the IMA(1,1) model then for an AR(4) model. For the IMA(1,1) model, the QLR test rejects at the 1% significance level, providing formal evidence of instability in the parameters of the UC model. The tests on individual parameters suggest that this instability appears in the permanent innovation variance but not the transitory innovation variance. For the AR(4) model, the null hypothesis of stability is rejected at the 5%, but not the 1%, significance level; the smaller *p*-value for the UC model is consistent with fewer degrees of freedom because fewer parameters are being tested for stability in the UC model than in the AR(4) model.

One measure of how important the permanent shocks are is to decompose the four-quarter ahead forecasts into three sources: errors in estimation of the current trend, that is, signal extraction (filtering) errors, forecast errors arising from currently unknown permanent disturbances over the next four quarters, and forecast errors arising from currently unknown transitory disturbances over the next four quarters. This decomposition is given in the final block of Table 3. In the first period, future trend disturbances are by far the largest source of four-quarter forecast errors, followed by filtering errors. The magnitude of both sources of error falls sharply from the first period to the second: the forecast error attributed to the trend disturbance falls by over 90%, and the forecast error variance arising from filtering error falls by 60%. Like the contribution of the trend shock itself, the decline in the contribution of the filtering error is a consequence of the decline in the volatility of the trend shock because the trend is less variable and therefore is estimated more precisely by the UC filter. The contribution of the transitory shocks remains small and is approximately unchanged between the two periods.

Historical precedents. The IMA(1,1) representation for inflation is not new. As mentioned in the introduction, Nelson and Schwert (1977) selected an IMA(1,1) model for monthly U.S. CPI inflation (identified, in the Box-Jenkins (1970) sense, by inspecting the autocorrelogram of π_t and $\Delta\pi_t$). Using a sample period of 1953m2 – 1971m7, they estimated an MA coefficient of .892 (their equation (4)). This monthly IMA(1,1) model temporally aggregates to the quarterly IMA(1,2) model[2],

$$\Delta\pi_t = (1 - .487\mathrm{B} - .158\mathrm{B}^2)a_t. \qquad \text{(NS77)} \qquad (7)$$

Schwert (1987) reports IMA(1,1) models for two monthly price indexes (the CPI and the PPI) and one quarterly index, the GNP deflator. Using data from 1947:I – 1985:IV, Schwert (1987, Table 6) estimated the MA coefficient for the GNP deflator to be .665.

A third historical reference is Barsky (1987), who uses Box-Jenkins identification methods to conclude that quarterly CPI inflation (third month of quarter aggregation) is well described by an IMA(1,1) model, with a MA coefficient that he estimates to be .46 over the 1960-1979 period (Barsky (1987, Table 2)).

We return to these historical estimates below.

3.3 Explaining the AO Results

All the univariate models considered so far, including the AO model, produce multistep forecasts that are linear in π_t, π_{t-1},…, so one way to compare these models is to compare their forecast functions, that is, their weights on π_t, π_{t-1},…. Figure 1 plots the forecast functions for four-quarter ahead forecasts computed using the AO model, an

[2] Let P_t^m be the monthly log price index. If $\Delta\ln(P_t^m)$ follows the monthly IMA(1,1) $\Delta^2\ln(P_t^m) = (1 - \theta^m\mathrm{B})a_t$, then the monthly latent quarterly inflation series follows $(1 - \mathrm{B}^3)^2 p_t^Q = (1 + \mathrm{B} + \mathrm{B}^2)^3(1 - \theta^m\mathrm{B})a_t$, which, when sampled every third month, corresponds to an IMA(1,2) at the quarterly frequency.

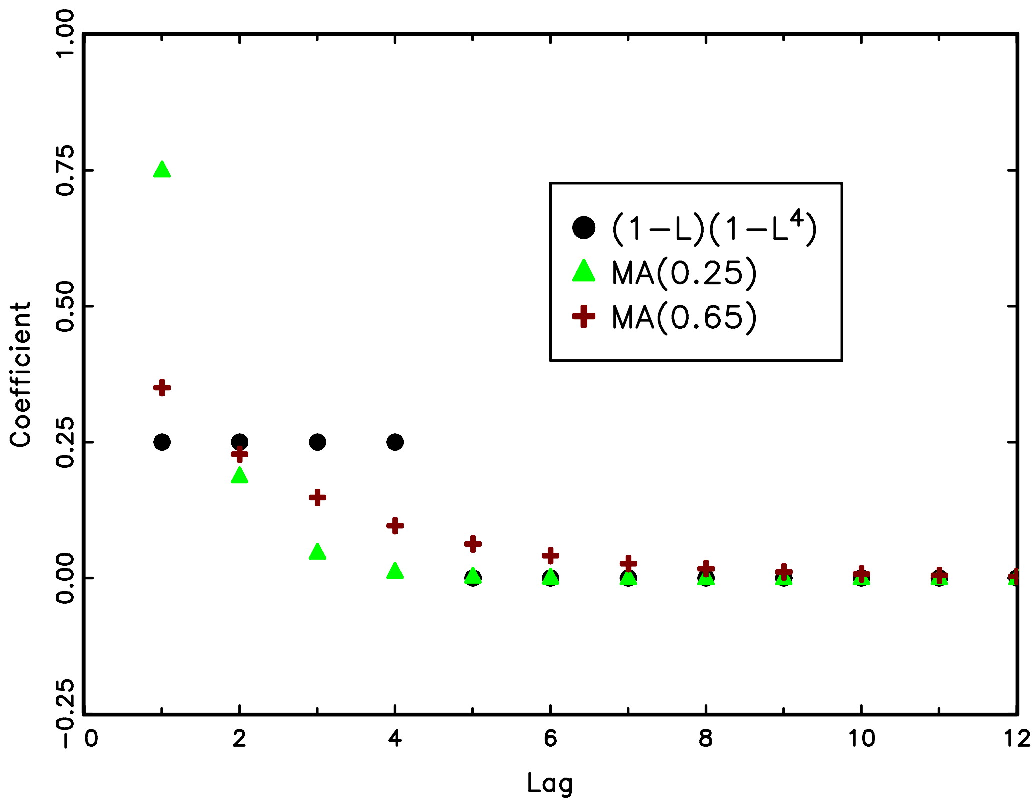


Figure 1. Implied forecast weights on lagged quarterly inflation for forecasts of four-quarter inflation computed using the Atkeson-Ohanian (2001) model $(1 - L)(1 - L^4)$ and using an IMA(1,1) model with $\theta = .25$ and $\theta = .65$.

IMA(1,1) model with $\theta = .25$, and an IMA(1,1) model with $\theta = .65$. (As discussed in Section 5, the value $\theta = .25$ closely approximates the value of θ estimated using the 1970-1983 sample for all four inflation series, and the value $\theta = .65$ closely approximates the estimates for all four series over the 1984-2004 period.) The AO forecast function weights the most recent four quarters of inflation evenly, whereas the IMA(1,1) forecast functions are geometrically declining. The AO and $\theta = .25$ forecast functions are quite different, and the resulting forecasts typically would be quite different. In contrast, the AO and $\theta = .65$ forecast function provides a closer approximation to the $\theta = .65$ forecast function, and one might expect the AO and $\theta = .65$ forecasts to be fairly close much of the time.

The changing coefficients in the IMA(1,1)/UC representation provide a concise arithmetic explanation for the performance of the AO forecasts evident in Table 1. Over the 1970-1983 period, during which the MA coefficient is small, the AO model would be expected to work poorly. During the later period, during which the MA coefficient is large, at the four-quarter horizon the AO model provides an approximation to the IMA(1,1) forecast and would be expected to work well. This approximation is also good at the eight-quarter horizon, but not at short horizons, so the AO model would be expected to work well at longer seasonal horizons, but not short horizons in the second period. This pattern matches that in Table 1.

4. Dating the Changes in the Inflation Process Using an Unobserved Components – Stochastic Volatility Model

The tests for parameter instability reported in Table 3 indicate that there have been statistically significant and economically large changes in the univariate inflation process. This section takes a closer look at when those changes occurred, and they are associated with continual parameter drift or discrete regime shifts. Is the IMA model with a single break in 1984 a satisfactory approximation to the inflation process, or have the changes been more subtle and evolutionary?

The model of this section is a generalization of the unobserved components model in which the variances of the permanent and transitory disturbances evolve randomly over time, that is, an unobserved components model with stochastic volatility (UC-SV). In the UC-SV model, logarithms of the variances of η_t and ε_t evolve as independent random walks. The UC-SV model is,

$$\pi_t = \tau_t + \eta_t, \quad \text{where } \eta_t = \sigma_{\eta,t}\zeta_{\eta,t} \tag{8}$$

$$\tau_t = \tau_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t = \sigma_{\varepsilon,t}\zeta_{\varepsilon,t} \tag{9}$$

$$\ln\sigma^2_{\eta,t} = \ln\sigma^2_{\eta,t-1} + \nu_{\eta,t} \tag{10}$$

$$\ln\sigma^2_{\varepsilon,t} = \ln\sigma^2_{\varepsilon,t-1} + \nu_{\varepsilon,t} \tag{11}$$

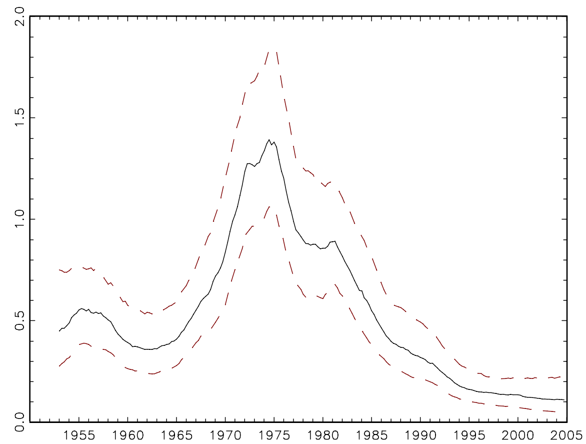
where $\zeta_t = (\zeta_{\eta,t},\ \zeta_{\varepsilon,t})$ is i.i.d. N(0, I_2), $\nu_t = (\nu_{\eta,t},\ \nu_{\varepsilon,t})$ is i.i.d. N(0, γI_2), and ζ_t and ν_t are independently distributed, and γ is a scalar parameter. Note that this model has only one parameter, γ, which controls the smoothness of the stochastic volatility process; γ can either be estimated or chosen *a-priori*.

Results. Figure 2 plots the smoothed estimates of $\sigma_{\eta,t}$ and $\sigma_{\varepsilon,t}$ from the UC-SV model, computed by Markov Chain Monte Carlo (MCMC) using a vague prior for the initial condition and $\gamma = 0.2$, for GDP inflation, using data from 1953:I – 2004:IV (the longer sample is used here to obtain estimates for the 1950s, and to facilitate comparisons with the Nelson-Schwert (1977) estimate, which was based on data starting in 1953).

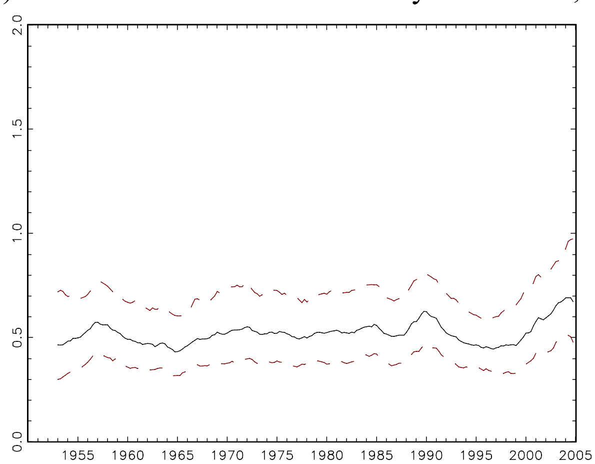
The estimates in Figure 2 show substantial movements over time in the standard deviation of the permanent component: the 1970s through 1983 was a period of high volatility, 1953 through the late 1960s or early 1970s and 1984-1990 were periods of moderate volatility of the permanent innovation, and the 1990s through 2004 have been period of low volatility of the permanent innovation. In contrast, there is little change in the estimates of the variance of the transitory innovation. The moving average coefficient of the implied instantaneous IMA(1,1) representation tracks inversely the smoothed estimates of $\sigma_{\varepsilon,t}$, being moderate (around .4) in the 1950s through late 1960s, small (less than .25) during the 1970s through 1983, higher in the late 1980s, and increasing further in the 1990s to a current estimate of approximately .85.

A UC-SV model with heavy-tailed volatility innovations. The UC-SV model specifies the log variances as following a Gaussian random walk. This imparts smoothness to the stochastic volatility, relative to a stochastic volatility process with heavier tails. If a regime shift model is a better description of the changes in volatility

(a) Standard deviation of permanent innovation, $\sigma_{\varepsilon,t}$



(b) Standard deviation of transitory innovation, $\sigma_{\eta,t}$



(c) Implied IMA(1,1) coefficient θ

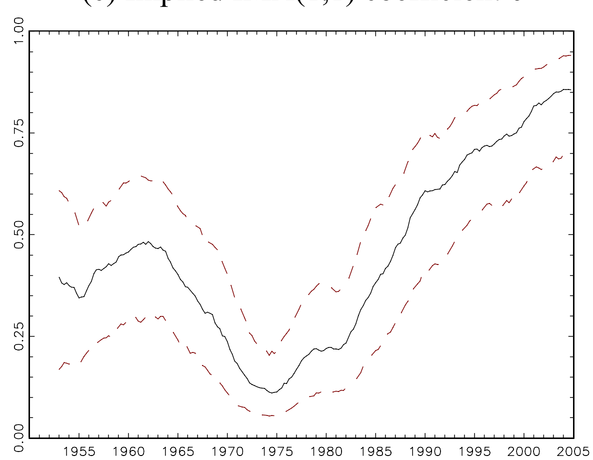


Figure 2. Estimates of the standard deviations of the permanent and transitory innovations, and of the implied IMA(1,1) coefficient, using the TC-SV(.2) model: 16.5%, 50%, and 83.5% quantiles of the posterior distributions, GDP deflator, 1953-2004

than a model, like the UC-SV model, with smooth parameter variation, then the UC-SV model (8) – (11) might miss the rapid changes in volatility associated with a shift in regimes.

To investigate this possibility, we modified the UC-SV so that the disturbances $\zeta_{\eta,t}$ and $\zeta_{\varepsilon,t}$ in (8) and (9) were drawn from a mixture of normal distributions, N(0,.1I) with probability .95 and N(0,.5I) with probability .05; this heavy-tailed mixture distribution introduces occasional large jumps. As it happens, the smoothed estimates of $\sigma_{\eta,t}$ and $\sigma_{\varepsilon,t}$ from the mixture-of-normals UC-SV model are qualitatively and quantitatively close to those from the normal-error UC-SV model. To conserve space, the mixture-of-normal UC-SV results are not presented, and for simplicity the only UC-SV model considered below is the normal-error model (8) – (11).

Relation to estimates in the literature. The series analyzed by Nelson and Schwert (1977), Schwert (1987), and Barsky (1987) differ from those analyzed here, and the Nelson-Schwert and Schwert estimates for CPI were computed using monthly data. Despite these differences, their estimated IMA(1,1) parameters are consistent with figures 3 and 4. The published estimates of the quarterly MA coefficient are .46 for 1960-1979 (Barsky (1987)) and .665 for 1949-1985 (Schwert (1987)). Comparing these estimates to averages in figure 2(c) over the corresponding time period indicates general agreement between the instantaneous MA coefficient estimated in figure 2(c) and the earlier estimates (the high estimate in Schwert (1987) seems to be driven in part by the pre-1953 data).

5. Pseudo Out-of-Sample Univariate Forecasts

This section examines whether unobserved components and moving average model with time-varying parameters could have produced useful real-time univariate forecasts.

5.1 Forecasting Models

The comparison considers two models studied in Table 1, the recursive AR(AIC) benchmark and the AO model, plus additional rolling and recursive models.

Recursive IMA(1,1) and AR(4). The models are estimated using an expanding sample starting in 1960.

Rolling AR(AIC), IMA(1,1), and AR(4). The models are estimated using a data window of forty quarters, concluding in the quarter of the forecast.

Nelson-Schwert (1977) (NS77). The NS77 model is the quarterly IMA(1,2) model (7) implied by temporal aggregation of Nelson and Schwert's (1977) monthly IMA(1,1) model.

UC-SV, γ = .2. Forecasts are computed using the UC-SV model with γ = .2. The UC-SV model is applied to data from 1960 through the forecast quarter to obtain filtered estimates of the trend component of inflation, which serves as the forecast for future values of inflation ($\pi_{t+h/t} = \tau_{t/t}$).

Fixed coefficient IMA(1,1). These are IMA(1,1) models with coefficients of .25 and .65. The coefficient of .25 approximately corresponds to the value in Table 2 for the period 1970 – 1983, and the coefficient of .65 approximately corresponds to the value for 1984 – 2004.

Multiperiod forecasts based on the IMA(1,1) and UC models are iterated, and multiperiod AR forecasts use the direct method (1).

The recursive and rolling models produce pseudo out-of-sample forecasts. The NS77 model produces a true out of sample forecast, since the coefficients were estimated using data through 1971. The UC-SV model with $\gamma = .2$ and the fixed-coefficient IMA(1,1) models are not pseudo out-of-sample models because their parameters (γ in the first instance, θ and σ_a in the second) were estimated (or, in the case of γ, calibrated) using the full data set, so in particular their parameters were not estimated by recursive or rolling methods.

5.2 Results

Table 4 summarizes the forecasting performance of the various models over the 1970-1983 and 1984-2004 periods for GDP inflation (entries are MSFEs, relative to the recursive AR(AIC), which is also the benchmark in Table 1). Figure 3 provides additional detail about the forecasting performance at different points in time by presenting a two-sided smoothed estimate of the relative MSFE (relative to the recursive AR(AIC) forecast), with exponential smoothing and a discount factor of .95 (end points are handled by simple truncation). Inspection of table 4 and figure 3 suggests five findings.

First, among the fixed-parameter models, the $\theta = .25$ model performs well in the first period, whereas the $\theta = .65$ model performs well in the second. This is consistent with the choice of these two parameter values as being approximately the MLEs of θ in the two periods. The UC-SV(.2) model evidently adapts well to the shifting parameter

Table 4. Pseudo Out-of-Sample Forecasting Performance of Additional Univariate Models: MSFEs, Relative to AR(AIC), GDP inflation

Model	h = 1	h = 2	h = 4	h = 8	h = 1	h = 2	h = 4	h = 8
Recursive forecasts								
AR(AIC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AO	1.95 (.69)	1.57 (.44)	1.06 (.20)	1.00 (.22)	1.22 (.29)	1.10 (.28)	**0.89** (.18)	**0.84** (.23)
MA(1)	**0.82** (.06)	0.83 (.08)	0.87 (.12)	0.89 (.17)	1.01 (.04)	1.03 (.06)	0.98 (.12)	0.89 (.20)
AR(4)	0.95 (.06)	1.08 (.07)	1.05 (.08)	0.93 (.09)	0.93 (.03)	0.96 (.04)	0.99 (.05)	0.94 (.07)
Rolling forecasts								
AR(AIC)	0.97 (.06)	1.05 (.07)	0.99 (.09)	**0.83** (.13)	0.95 (.10)	1.08 (.13)	1.17 (.20)	1.18 (.37)
AR(4)	0.98 (.07)	1.15 (.09)	1.06 (.11)	0.94 (.15)	**0.92** (.09)	0.95 (.13)	1.06 (.20)	1.04 (.37)
MA(1)	**0.82** (.06)	**0.82** (.08)	**0.86** (.12)	0.88 (.18)	0.99 (.08)	0.98 (.10)	0.93 (.14)	0.87 (.21)
Nelson-Schwert								
NS77 MA(2)	0.88 (.23)	0.91 (.27)	0.95 (.26)	0.89 (.27)	0.93 (.09)	**0.91** (.14)	0.92 (.19)	**0.84** (.25)
Fixed-parameter models								
UC-SV, $\gamma = 0.2$	0.77	0.79	0.82	0.88	0.96	0.94	0.90	0.83
MA(1) $\theta = 0.25$	0.79 (.07)	0.80 (.08)	0.82 (.12)	0.87 (.17)	1.05 (.07)	1.11 (.10)	1.05 (.15)	0.93 (.22)
MA(1) $\theta = 0.65$	0.97 (.26)	0.94 (.25)	0.96 (.23)	0.90 (.24)	0.90 (.10)	0.87 (.14)	0.89 (.18)	0.82 (.23)

Notes to Table 4: Entries are MSFEs, relative to the recursively estimated AR(AIC). Bold entries are the smallest relative MSFE for the indicated series/period/horizon, among the out-of-sample (NS77) and pseudo out-of-sample forecasts. The fixed-parameter models do not generate pseudo out-of-sample forecasts because their parameters are not estimated using recursive or rolling samples. Standard errors for the relative MSFEs for the univariate models are given in parentheses. The standard errors were computed by parametric bootstrap with 5,000 draws. Specifically, synthetic quarterly data from 1959:I – 2004:IV were generated from the estimated UC-SV model using as the volatility parameters at each date the median estimate in Figure 2 (a and b), and the various univariate forecasting models were estimated and implemented using each artificial data draw, and the standard deviation of the synthetic counterparts to the entry in this table is the parametric bootstrap standard error.

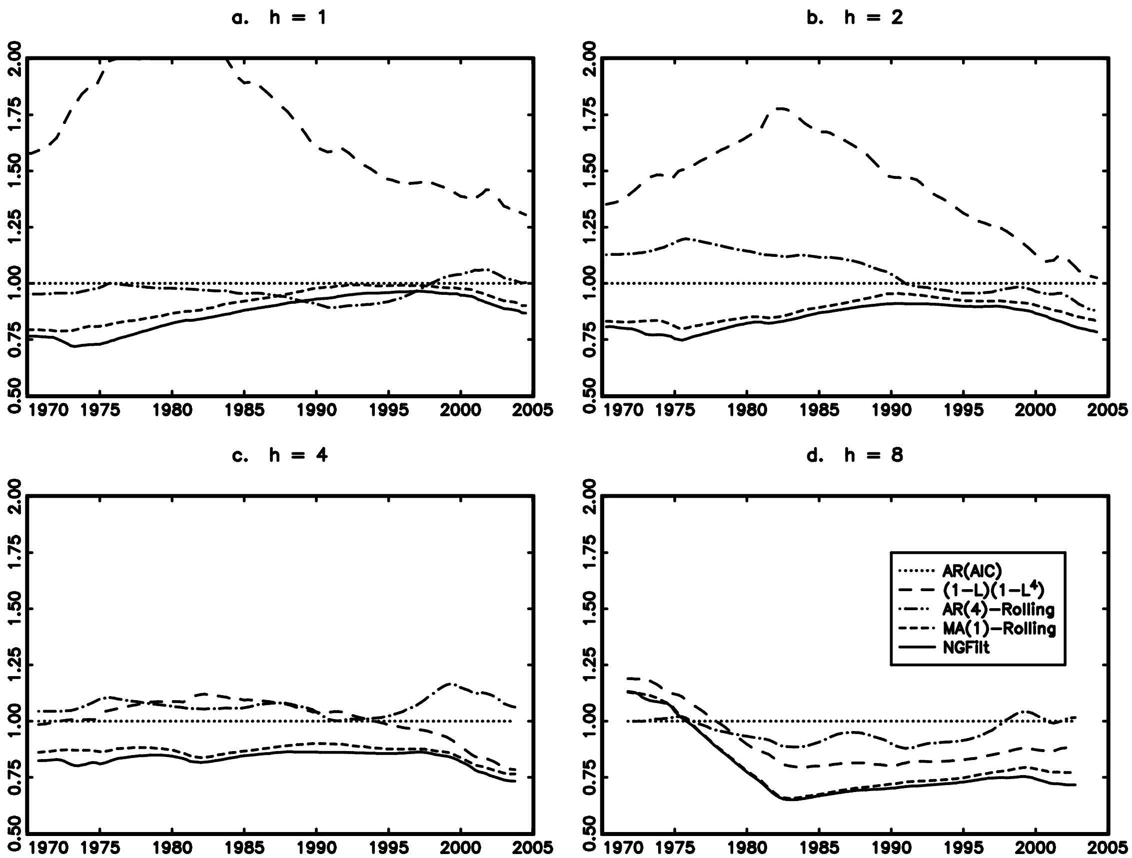


Figure 3. Smoothed relative mean squared forecast errors of various forecasts, relative to the recursive AR(AIC) benchmark: GDP deflator

values and produces forecasts that rival those of the $\theta = .25$ model in the first period, and those of the $\theta = .65$ model in the second. In this sense, among UC (IMA(1,1)) models, the UC-SV(.2) model can be thought of as providing an approximate bound on the forecasting performance of the pseudo out-of-sample forecasts.

Second, among the pseudo out-of-sample forecasts, the rolling IMA(1,1) forecast performs very well – nearly as well as the UC-SV model – in both periods. Closer inspection of the forecast errors indicates that the primary source of the improvement of the AO model over the rolling IMA(1,1) model during the second period occurs during the late 1980s, a period of sharp change in the MA coefficient during which the rolling forecast took time to adapt.

Third, the AR models do not forecast as well as the rolling IMA and UC/SV models. Eliminating the AIC lag selection by using an AR(4) improves performance in some but not all sample/horizon combinations; this is also true if AIC lag selection is replaced by BIC (unreported results).

Fourth, the NS77 forecast is truly an out-of-sample forecast, and its performance is in the post-1984 sample at all horizons is remarkably good.

Fifth, although the differences in the forecasting performance between these models is large in an economic and practical sense, this conclusion should be tempered by recognizing that there is a great deal of sampling uncertainty. Table 4 reports standard errors, which were computed by parametric bootstrap using the estimated time-varying UC model (using the median estimates of the time-varying variances) as the data generating process. As emphasized by Clark and McCracken (2005), there is considerable sampling variability in the relative MSFEs, which are often within one standard deviation of 1.00 (note however that the relative MSFEs for nested models do not have a normal distribution). Given these large standard errors, we suggest the following interpretation. The in-sample analysis of Section 4 provides strong evidence, which includes formal hypothesis tests, that there has been time variation in the inflation process and that this time variation is well described by the UC model with time-varying variances. That analysis suggests that a rolling MA(1) model will improve upon other univariate linear forecasting models. As measured by the relative MSFEs in Table 4, this prediction is borne out, and the improvement is economically large but imprecisely estimated.

5.3 Results for Other Price Indexes

Table 5 summarizes results for three other price indexes (PCE-core, PCE-all, and CPI). For PCE-core and PCE-all, in virtually all regards the results are quantitatively and qualitatively similar to those in Tables 3 and 4, so none of the main findings discussed so far hinge on using GDP inflation instead of either of these other price indexes. All the 90% confidence intervals for the largest AR root include one; both PCE series are well modeled using the TVP-UC model; and the QLR statistic rejects parameter stability in the UC model. The forecasts from the rolling IMA(1,1) are in most cases the best or nearly the best among the forecasts considered; the main exception to this is that the AO model outforecasts the rolling IMA(1,1) model in the second period, due mainly to the slow adaptation of the rolling IMA(1,1) to the rapidly changing parameters in the mid-1980s. The main difference between the results for the CPI and for

Table 5. Estimated Models and Forecasting Summary for Other Price Indexes

	PCE-core		PCE-all		CPI	
	1960:I – 1983:IV	**1984:I – 2004:IV**	**1960:I – 1983:IV**	**1984:I – 2004:IV**	**1960:I – 1983:IV**	**1984:I – 2004:IV**
(a) IMA parameters: $\Delta\pi_t = (1 - \theta B)a_t$						
θ	0.252 (.063)	0.677 (.094)	0.249 (.094)	0.688 (.088)	0.301 (0.085)	0.695 (0.085)
σ_a	1.053 (.058)	0.604 (.055)	1.273 (.080)	0.966 (.073)	1.769 (0.097)	1.333 (0.080)
(b) UC parameters						
σ_ε	0.787 (.079)	0.195 (.063)	0.957 (.121)	0.301 (.093)	1.235 (0.156)	0.407 (0.110)
σ_η	0.529 (.074)	0.497 (.051)	0.635 (.080)	0.801 (.072)	0.971 (0.153)	1.111 (0.107)
(c) p-values of Wald Tests of IMA(1,1) versus:						
ARIMA(1,1,1)	0.32	0.33	0.98	0.91	0.72	0.01
IMA(1,4)	0.66	0.73	.004	0.46	<.001	.002
(d) ARIMA(1,0,1) parameters: $(1 - \phi B)\pi_t = (1 - \theta B)a_t$						
ϕ	0.990 (.017)	0.992 (.008)	0.986 (.019)	0.992 (.013)	0.982 (.021)	0.986 (.015)
θ	−0.243 (.105)	−0.679 (.083)	−0.240 (.104)	−0.687 (.083)	–0.301 (.102)	–0.693 (.084)
(e) 90% confidence interval for largest AR root						
	0.889 - 1.030	0.913 - 1.040	0.859 - 1.026	0.834 - 1.029	0.856- 1.025	0.721- 1.002
(f) Tests for parameter stability						
t-statistic for $\sigma_{\varepsilon,70\text{-}83} = \sigma_{\varepsilon,84\text{-}04}$ (*p*-value)	–	–5.89 (<.001)	–	–4.31 (<.001)	–	–4.33 (<.001)
t-statistic for $\sigma_{\eta,70\text{-}83} = \sigma_{\eta,84\text{-}04}$ (*p*-value)	–	–0.35 (.727)	–	1.08 (.278)	–	0.753 (0.452)
QLR: UC model (*p*-value)	–	24.84 (<.01)	–	17.69 (<.01)	–	22.33 (<.01)
QLR: AR(4) (*p*-value)	–	4.77 (.01)	–	5.01 (<.01)	–	5.91 (<.01)
(g) 4-quarter ahead pseudo out-of-sample relative MSFEs (recursive AR(AIC) = 1.00)						
	1970:I – 1983:IV	**1984:I – 2004:IV**	**1970:I – 1983:IV**	**1984:I – 2004:IV**	**1970:I – 1983:IV**	**1984:I – 2004:IV**
AO	1.14	**0.71**	1.13	0.74	1.11	0.78
recursive IMA(1,1)	0.89	0.89	0.84	0.90	**0.87**	0.90
recursive AR(4)	1.13	0.97	1.05	1.00	0.98	1.05
rolling AR(AIC)	0.99	1.05	0.96	0.94	0.97	0.91
rolling AR(4)	1.17	1.05	1.08	0.98	1.01	0.95
rolling IMA(1,1)	**0.87**	0.81	**0.82**	0.80	0.87	0.85
NS77 MA(2)	0.99	0.75	1.01	**0.73**	1.04	**0.70**
UC-SV, $\gamma = 0.2$	0.88	0.78	0.80	0.87	0.86	1.06

Notes to Table 5: For part (e), see the notes to Table 2, for part (g), see the notes to Table 4, and for the other parts, see the notes to Table 3.

the other inflation series is that the quarterly CPI does not seem to be well modeled as a IMA(1,1) (the IMA(1,1) model is rejected against the IMA(1,4) alternative in both samples). One issue with interpreting the results for the CPI, especially model stability results, is that the construction of the CPI is not consistent over time, in particular the treatment of housing has changed. Perhaps more important, however, are issues of temporal aggregation. The measurement procedure for the GDP price index, the PCE price index, and the CPI are all different and paying attention to the measurement details could yield a different method of temporal aggregation than the stylized model used in footnote 2. The fact that the temporally aggregated NS77 forecasts (see (7)) are excellent for the CPI is consistent with the monthly (but not quarterly) CPI following an IMA(1,1), and suggest that closer attention to temporal aggregation issues, especially for the CPI, is warranted.

Additional results for PCE-core, PCE-all, and CPI inflation are presented in the Appendix. The breakdown of the ADL Philips curve forecasts seen in Table 1 is observed using these other inflation measures. PCE-core and PCE-all inflation are well described by a IMA(1,1) model with time-varying coefficients: for both indexes, the first (and only the first) autocorrelation of inflation is statistically significant from zero, a nonparametric estimate of the spectrum of $\Delta\pi_t$ approximately has the shape implied by a MA(1) model. The autocorrelations of CPI inflation suggest that a different model, perhaps an IMA(1,2), is more appropriate, which (as discussed above) could be a consequence of temporal aggregation. Still, the time paths of the standard deviations of the UC-SV model for all three indexes, including CPI inflation, are qualitatively similar to those for GDP inflation. In addition, the rolling IMA(1,1) model forecasts well for all these series, including the CPI, at all horizons in both sample periods. The NS77 MA(2) forecasts remarkably well in the second sample period: of sixteen possible cases (four horizons and four series), in the second sample period the NS77 model produces the best forecasts in eleven cases.

5.4 Reconciling the AR and MA Results

Long-order autoregressions provide arbitrarily good approximations to moving average processes in population, so any explanation of the difference between the rolling AR and rolling MA forecasts must appeal to finite-sample differences between the two estimated models. There are three such differences that are relevant and that can reconcile the AR and the MA results. First, in the second sample, θ is larger which implies that more distant autoregressive coefficients will be larger in absolute value, increasing the truncation bias of a finite-lag AR. Second, for this reason, the AIC will tend to call for longer AR lags, however this introduces greater estimation variance in the AR. Third, the implied population AR coefficients are larger in absolute value when θ is larger, and the larger population AR coefficient means that its estimator will have a larger bias towards zero. Each of these three finite-sample effects work towards making the AR model a worse approximation in the second sample than in the first.

5.5 Summary

The results in Sections 3 – 5 provide a simple picture of the evolution of the inflation process. The variability of the stochastic trend in inflation increased in magnitude during the 1970s through 1983, then fell significantly, both in statistical and economic terms. Although the stochastic trend component of inflation diminished in importance, it remains nonzero (said differently, confidence intervals for the largest AR root continue to include one). Moreover, because the variance of the permanent (but not transitory) component fell dramatically, the variance of $\Delta\pi_t$ has fallen. Because the smaller permanent component variance corresponds to a larger MA coefficient, the declining importance of the stochastic trend in inflation explains both the good performance of the AO forecast in the second sample at long horizons, and its poor performance at short horizons and in the first sample. An important piece of evidence supporting this interpretation is that the rolling IMA(1,1) model produces forecasts that have the lowest MSFE, or nearly so, at all horizons for all four inflation series among the recursive and rolling univariate models, including the AO model.

The declining importance of the permanent component (equivalently, the increase in the MA coefficient) implies that an AR approximation needs more lags and larger coefficients, two features that work towards increased sampling variability and greater finite-sample bias of AR forecasts. This is consistent with the better performance of the rolling MA forecasts, compared to the rolling AR forecasts, since the mid-1980s.

6. Implications for Multivariate Forecasts

The foregoing univariate analysis has three implications for the specification of conventional ADL Phillips curves of the form (3):

1. The coefficients on lags of $\Delta\pi_t$ will, to a first approximation, decline exponentially as would be implied by inverting a MA(1) (this implication holds exactly if lagged x_t is uncorrelated with lagged $\Delta\pi_t$). The restricted one-step ahead specification has a rational lag specification for these coefficients:

$$\Delta\pi_t = \mu + \beta x_{t-1} + \delta(B)\Delta x_{t-1} + \alpha(B)\Delta\pi_{t-1} + u_t, \text{ where } \alpha(B) = -\psi(1-\psi B)^{-1}. \quad (12)$$

   If the coefficients β and $\delta(B)$ are zero, the rational lag parameter ψ would equal the MA parameter θ.

2. There will be instability in the coefficients on lags of $\Delta\pi_t$ over the sample, with the coefficients being larger in absolute value since the mid-1980s than during the 1970s and early 1980s.

3. Failure to allow for time variation in the coefficients on lagged $\Delta\pi_t$ could lead to an apparent shift in the coefficients on x_t; in principle the Phillips relation could be stable once one allows for changes in the coefficients on lagged $\Delta\pi_t$.

This section examines these three implications. The analysis is in-sample (not forecasting) and focuses on specifications with the unemployment gap and the output gap, where both one- and two-sided gaps are considered. The two-sided gap form is useful for historical analysis and assessing coefficient stability. The one-sided gap form is useful for assessing forecasting content.

Table 6 reports estimates of the one-step ahead model using the unrestricted ADL(4,4) specification (3) (first two columns) and the restricted specification (12) (final two columns), where the activity variable x_t is the unemployment gap (one- and two-sided, parts a and b) and the output gap (one- and two-sided, parts c and d). Table 6 suggests five conclusions.

First, the coefficients on lags of $\Delta\pi_t$ in the unrestricted model are qualitatively and quantitatively consistent with the exponential decline in the restricted model. These results support implication 1 above. Other researchers have stressed the importance of using long restricted specifications for the lags on $\Delta\pi_t$ in Phillips curves. Gordon's (1998 and earlier) specifications use lagged annual inflation rates, in effect imposing a step function specification on the lag shape for quarterly inflation. Brayton, Roberts, and Williams (1999) find that polynomial distributed lag specifications with long lags (up to 25 quarters) fit better than models with shorter unrestricted lags. The results in Table 6 support the use of restricted, long-lag specifications. However the functional form in Table 6 differs from these other authors, in that it is a rational lag specification, not a step function or a polynomial distributed lag.

Second, consistent with implication 2 above, the coefficients on lags of $\Delta\pi_t$ increase in absolute value from the pre-1984 to the post-1984 samples. As in the IMA(1,1) model, the rational lag parameter ψ increases from the first to the second sample, indeed, the estimated parameter ψ in the multivariate model is quantitatively close to the estimated MA parameter θ in each sample, for both activity gaps.

Table 6. ADL and Restricted ADL Models of GDP Inflation:

$$\Delta\pi_t = \mu + \beta x_{t-1} + \delta(B)\Delta x_{t-1} + \alpha(B)\Delta\pi_{t-1} + u_t$$

a. $x_t = ugap^{1\text{-}sided}$

	α(B) unrestricted, 4 lags, OLS		$\alpha(B) = -\psi(1-\psi L)^{-1}$, NLLS	
	1960:I − 1983:IV	**1984:I − 2004:IV**	**1960:I − 1983:IV**	**1984:I − 2004:IV**
(i) Coefficients (standard errors)				
x_{t-1}	–0.367 (.258)	–0.100 (.148)	−.330 (.262)	−0.143 (.087)
$\Delta\pi_{t-1}$	–0.309 (.128)	–0.666 (.123)	*−0.366*	*−0.665*
$\Delta\pi_{t-2}$	–0.234 (.113)	–0.503 (.121)	*−0.134*	*−0.442*
$\Delta\pi_{t-3}$	–0.144 (.132)	–0.339 (.124)	*−0.049*	*−0.294*
$\Delta\pi_{t-4}$	0.017 (.102)	0.116 (.104)	*−0.018*	*−0.200*
ψ	–	–	0.366 (.111)	0.665 (.082)
SER	1.222	0.730	1.219	0.758
$R^2_{x\|\text{lags of }\Delta\pi}$	.103	.038	.111	.029
Granger causality F-test (*p*-value)	3.45 (0.01)	1.23 (0.29)	3.72 (< .01)	1.01 (0.40)
(ii) Chow test for a break in coefficients on (p-value):				
x_{t-1}		0.37		0.54
Δx_{t-1}		0.48		0.78
$\Delta\pi_{t-1}$		0.16		0.03
all coefficients		0.10		0.05

b. $x_t = ugap^{2\text{-}side}$

	α(B) unrestricted, 4 lags, OLS		$\alpha(B) = -\psi(1-\psi L)^{-1}$, NLLS	
	1960:I − 1983:IV	**1984:I − 2004:IV**	**1960:I − 1983:IV**	**1984:I − 2004:IV**
(i) Coefficients (standard errors)				
x_{t-1}	−0.406 (.151)	−0.183 (.092)	−0.438 (.157)	−0.288 (.119)
$\Delta\pi_{t-1}$	−0.366 (.128)	−0.742 (.110)	*−0.467*	*−0.717*
$\Delta\pi_{t-2}$	−0.290 (.111)	−0.558 (.119)	*−0.218*	*−0.514*
$\Delta\pi_{t-3}$	−0.180 (.117)	−0.366 (.128)	*−0.102*	*−0.369*
$\Delta\pi_{t-4}$	0.041 (.098)	0.070 (.106)	*−0.048*	*−0.264*
ψ			0.467 (.106)	0.717 (.075)
SER	1.185	.679	1.188	0.701
$R^2_{x\|\text{lags of }\Delta\pi}$	.152	.118	.152	.125
Granger causality F-test (*p*-value)	3.67 (<.01)	4.88 (<.01)	3.31 (.01)	5.34 (< .01)
(ii) Chow test for a break in coefficients on (p-value):				
x_{t-1}		0.21		0.44
Δx_{t-1}		0.27		0.57
$\Delta\pi_{t-1}$		0.18		0.05
all coefficients		0.04		0.02

Table 6, ctd.

c. $x_t = ygap^{1-sided}$

	α(B) unrestricted, 4 lags, OLS		α(B) = $-\psi(1-\psi L)^{-1}$, NLLS	
	1960:I − 1983:IV	1984:I − 2004:IV	1960:I − 1983:IV	1984:I − 2004:IV
(i) Coefficients (standard errors)				
x_{t-1}	0.186 (.087)	0.109 (.076)	0.174 (.088)	0.136 (.081)
$\Delta\pi_{t-1}$	−0.297 (.123)	−0.634 (.123)	*−0.368*	*−0.659*
$\Delta\pi_{t-2}$	−0.254 (.111)	−0.474 (.128)	*−0.135*	*−0.434*
$\Delta\pi_{t-3}$	−0.114 (.133)	−0.329 (.131)	*−0.050*	*−0.286*
$\Delta\pi_{t-4}$	0.054 (.108)	0.124 (.110)	*−0.018*	*−0.189*
ψ			0.368 (.110)	0.659 (.0978)
SER	1.253	0.723	1.256	0.754
$R^2_{x\vert\text{lags of }\Delta\pi}$	.061	.050	.059	.037
Granger causality F-test (*p*-value)	1.40 (0.23)	1.36 (0.24)	1.36 (0.25)	1.12 (0.35)
(ii) Chow test for a break in coefficients on (p-value):				
x_{t-1}		0.51		0.75
Δx_{t-1}		0.48		0.79
$\Delta\pi_{t-1}$		0.24		0.05
all coefficients		0.09		0.06

d. $x_t = ygap^{2-sided}$

	α(B) unrestricted, 4 lags, OLS		α(B) = $-\psi(1-\psi L)^{-1}$, NLLS	
	1960:I − 1983:IV	1984:I − 2004:IV	1960:I − 1983:IV	1984:I − 2004:IV
(i) Coefficients (standard errors)				
x_{t-1}	0.159 (.059)	0.068 (.052)	0.149 (.060)	0.106 (.062)
$\Delta\pi_{t-1}$	−0.326 (.123)	−0.649 (.120)	*−0.427*	*−0.693*
$\Delta\pi_{t-2}$	−0.308 (.113)	−0.517 (.124)	*−0.182*	*−0.480*
$\Delta\pi_{t-3}$	−0.155 (.135)	−0.355 (.126)	*−0.078*	*−0.333*
$\Delta\pi_{t-4}$	0.009 (.109)	0.092 (.109)	*−0.033*	*−0.231*
ψ			0.427 (.108)	0.693 (.088)
SER	1.233	0.714	1.240	0.739
$R^2_{x\vert\text{lags of }\Delta\pi}$	.088	.064	.082	.063
Granger causality F-test (*p*-value)	2.23 (.06)	2.07 (.08)	2.20 (0.07)	1.85 (0.12)
(ii) Chow test for a break in coefficients on (p-value):				
x_{t-1}		0.25		0.62
Δx_{t-1}		0.51		0.83
$\Delta\pi_{t-1}$		0.23		0.06
all coefficients		0.09		0.11

Notes to Table 6: Entries in block (i) are the coefficients in the autoregressive distributed Phillips curve given in the table header, for the sample period given in the column header, for the unrestricted model (first two columns) and the restricted model (final two columns), where the restricted model imposes the MA(1) (rational lag) functional form on the parameters. Standard errors are in parentheses. *SER* is the standard error of the regression and $R^2_{x|\text{lags of }\Delta\pi}$ is the regression partial R^2 for x_{t-1}, Δx_{t-1}, and lags of Δx_{t-1}. The Granger causality test statistic tests the joint significance of the coefficients on x_{t-1}, Δx_{t-1}, and lags of Δx_{t-1}. Block (ii) reports *p*-values for heteroskedasticity-robust Chow tests of a break in the indicated coefficient or set of coefficients, with a break date of 1984:I. The activity variable x_t analyzed in each part of the table (a-d) is given in the part header.

Third, in the restricted rational distributed lag model, the Chow test for a break in 1984 in the coefficients on lags of $\Delta\pi_t$ rejects at the 5% level in three of the four specifications (and in all four at the 10% level). However, in the unrestricted ADL model, the same Chow test does not reject at the 10% level in any of the four specifications. The greater degrees of freedom of the Chow test in the unrestricted specification evidently reduces power by enough to mask the changes in those coefficients.

Fourth, the evidence on implication 3 – whether the Phillips curve is stable after allowing for changes in the coefficients on lagged $\Delta\pi_t$ – is mixed. There is some evidence that imposing the rational lag specification with a changing parameter ψ leads to a more stable coefficient on the activity gap x_t. For example, in the one-sided unemployment gap specification, the coefficient on x_t fell by 73% in absolute value between the two samples in the unrestricted specification (from –.367 to –.100), but only fell by 56% in the restricted specification. Although these changes are large in an economic sense, they are imprecisely estimated, and the hypothesis that the coefficient on x_{t-1} is constant across the two samples is not rejected in any of the eight cases considered in Table 6.

Fifth, the marginal explanatory content of the activity variables dropped substantially from the first to the second sample, both in the unrestricted and restricted specifications. Neither 1-sided gap variable produces a significant Granger causality test statistic in the second sample, and the partial R^2 of the 1-sided gap variables is quite small in the second period, less than .04 for both restricted specifications. Although these results are in-sample, consider only two gap variables, and are only for a one-step ahead specification, they suggest that it could be challenging to use the time variation found in the univariate analysis to develop useful activity-based forecasts.

Additional evidence of some stability in the relationship between four-quarter GDP inflation and gap measures is presented in Figure 4, which plots the four-quarter ahead prediction error from the UC-SV model ($\gamma = .2$) and the two-sided unemployment gap. The slope of this bivariate regression is less in the second period than in the first, and the $\bar{R}^2$ falls by roughly one-half between the two periods. Still, for GDP inflation, the slope is statistically significant (and negative for the unemployment gap, positive for the output gap) in both periods. Also, in results not tabulated here, the implied coherence between the activity variables and changes of inflation at business cycle frequencies is large and relative stable across periods, typically being in the range 0.5 – 0.6. This stable coherence is consistent with the positive and stable association found at business cycle frequencies by Harvey, Trimbur, and van Dijk (2005, figures 16 and 17), who used a bivariate unobserved components (trend-cycle) model with different, but possibly correlated, real and nominal cyclical components.

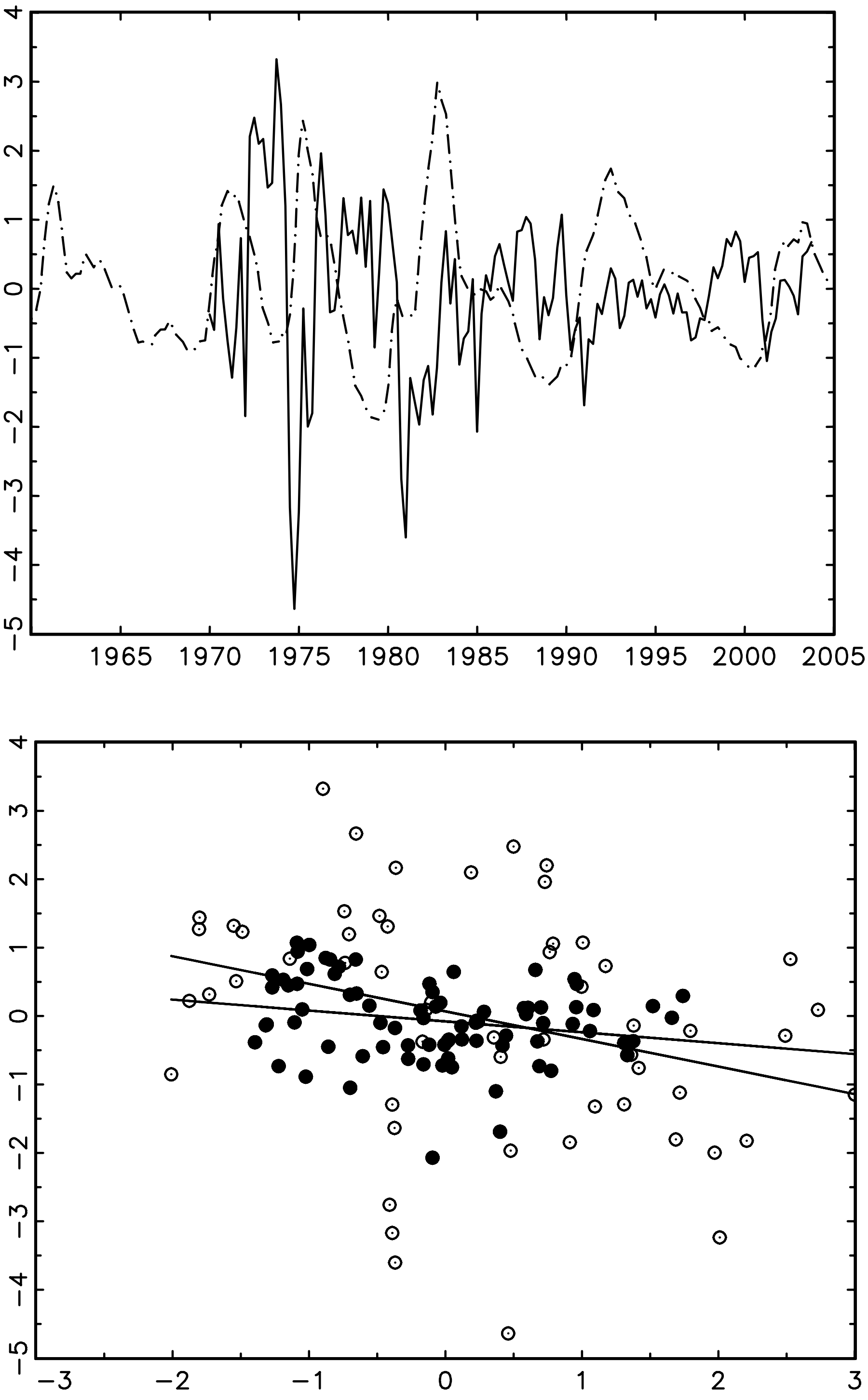


Figure 4. Time series plot (upper) and scatterplot of residual (lower) from UC-SV model of GDP inflation vs. the 2-sided unemployment gap. Scatterplot: open circles, 1970–1983; filled circles, 1984–2004.

7. Discussion and Conclusions

This paper has explored some implications of the changing univariate inflation process for multivariate activity-based inflation forecasting, but more work remains. These results present the hope that, with suitable modification using a rational lag specification like (12), the outlook for forecasting using backward-looking Phillips curves might be less gloomy than the results in AO and in Section 2 would lead one to believe. This said, it is not straightforward to turn the increased stability in two-sided gap models into reliable one-sided gap forecasting specifications. Still, there is some evidence that real-time forecasts have provided improvements upon the best univariate models. Kohn (2005) reports large reductions in true (not pseudo) mean squared errors of real-time Fed staff forecasts of quarterly CPI inflation, relative to the AO model, over 1984-2000, and Ang, Bekaert, and Wei (2005) find that true out-of-sample survey forecasts (median Michigan or Livingston forecasts) outperform a large number of pseudo out-of-sample univariate and multivariate time series competitors. Both the Board staff forecast and the survey forecasts are combination forecasts, pooled over judgmental and model-based forecasts, and both presumably incorporate considerably more information than are present in the simple activity-based forecasts examined here. Perhaps further attempts to develop competitive time series forecasts could profit from pursuing systematically those features that have proven successful in these survey forecasts.

One thing this paper has not done is to attempt to link the changing time series properties of inflation to more fundamental changes in the economy. The obvious explanation is that these changes stem from changes in the conduct of monetary policy in the post-1984 era, moving from a reactive to a forward-looking stance (see for example the recent discussion in Estrella (2005), who explains the post-80s failure of the term structure to have predictive content for inflation in terms of changes in Fed policy). But obvious explanations are not always the right ones, and there are other possible reasons for the decrease in the variability of the permanent component of inflation. To a considerable extent, these other possibilities are similar to the ones raised in the context of the discussion of the great moderation, including changes in the structure of the real economy, the deepening of financial markets, and possible changes in the nature of the structural shocks hitting the economy. We do not attempt to sort through these explanations here, but simply raise them to point out that the question of deeper causes for these changes merits further discussion.

References

Ang, A., G. Bekaert, and M. Wei (2005), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" NBER Working Paper no. 11538.

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61:821-856.

Atkeson, A. and L.E. Ohanian (2001), "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review* 25(1):2-11.

Barsky, R.B. (1987), "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics* 19:3-24.

Benati, L. and H. Mumtaz (2005) "The 'Great Stability' in the U.K.: Good Policy or Good Luck?" manuscript, Research Department, Bank of England.

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis*. San Francisco.

Brayton, F., J.M. Roberts, and J.C. Williams (1999), "What Happened to the Phillips Curve?," FEDS Working Paper 1999-49.

Clark, T.E. and M.W. McCracken (2005), "The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence," forthcoming, *Journal of Money, Credit, and Banking*.

Dossche, M. and G. Everaert (2005), "Measuring Inflation Persistence: A Structural Time Series Approach," ECB Working Paper 495.

Estrella, A. (2005), "Why Does the Yield Curve Predict Output and Inflation," *The Economic Journal* 115:722-744.

Fisher, J.D.M., C.T. Liu, and R. Zhou (2002), "When Can We Forecast Inflation?" *Federal Reserve Bank of Chicago Economic Perspectives* 1Q/2002, 30-42.

Gordon, R.J (1998), "Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU," *Brookings Papers on Economic Activity* 1998(2):197-333.

Harvey, A.C., T.M. Trimbur, and H.K. van Dijk (2005), "Trends and Cycles in Economic Time Series: A Bayesian Approach," manuscript, Cambridge University, UK.

Kohn, D.L. (2005), "Modeling Inflation: A Policymaker's Perspective," speech to the International Research Forum on Monetary Policy, Frankfurt, Germany, May 20, 2005, available at http://www.federalreserve.gov/boarddocs/speeches/2005/20050520/default.htm.

Leigh, D. (2005), "Estimating the Implicit Inflation Target: An Application to U.S. Monetary Policy," IMF Working Paper 05/77.

Li, Hong (2005), "Testing Alternative Models of Price Adjustment," mimeo, Brandeis University.

Mavroeidis, S. (2005). "Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve," *Journal of Money, Credit, and Banking* 37:421-448.

Nelson, C.R. and G.W. Schwert (1977), "Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Rate of Interest Is Constant," *American Economic Review* 67:478-486.

Orphanides, A. and S. van Norden (2005), "The Reliability of Inflation Forecast Based on Output Gap Estimates in Real Time," *Journal of Money, Credit, and Banking* 37, 583-600.

Pivetta, F. and R. Reis (2004), "The Persistence of Inflation in the United States," manuscript, Princeton University.

Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series," *Journal of Monetary Economics*, 28:435-460.

Stock, J.H., and M.W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics* 44:293-335.

Stock, J.H. and M.W. Watson (2002), "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual* 2002, 159–218.

Stock, J.H., and M.W. Watson (2003), "Forecasting output and inflation: The role of asset prices," *Journal of Economic Literature* 41:788-829.

Tulip, P. (2005), "Has Output Become More Predictable? Changes in Greenbook Forecast Accuracy," FEDS no. 2005-31.

Appendix

The tables and figures in this appendix provide results for inflation as measured by the PCE-core, PCE-all, and CPI. The results complement those for GDP inflation presented in the text.

Figure A.1 plots the estimated spectrum of the change in PCE-all inflation. Two estimates are reported: a nonparametric estimator (smoothed periodogram) and a parametric IMA(1,1) estimator. The parametric estimate looks like a smoothed version of the nonparametric estimate, suggesting that the IMA(1,1) model fits the data reasonably well. Relative to the first period, the spectrum in the second period is lower in magnitude – this reflects the reduction in volatility between the two periods – and has more power at high than at low frequencies. Closer inspection reveals that the shape of the spectrum has changed, as well as its level, with the second period having relatively more power at higher frequencies than in the first. This is consistent with the more negative first autocorrelation in the second period than in the first.

Table A.1
Pseudo Out-of-Sample Forecast Results for the PCE Deflator

A. PCE Deflator (Core)

	1970:I – 1983:IV				1984:I – 2004:IV				$\frac{RMSFE_{84-04}^{h=4}}{RMSFE_{70-83}^{h=4}}$
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	
AR(AIC) RMSFE	1.37	1.46	1.66	2.11	0.68	0.60	0.56	0.62	
Relative MSFEs									
AR(AIC)	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	1.00	0.33
AO	2.55	1.90	1.14	0.99	1.59	1.11	**0.71**	**0.59**	0.26
PC-u u	1.08	0.99	0.94	0.60	1.03	1.04	1.32	1.58	0.40
PC Δu	**0.98**	**0.90**	**0.83**	0.65	1.06	1.14	1.45	1.95	0.44
PC $u-\bar{u}^{1-Sided}$	1.07	0.97	1.16	1.00	1.07	1.16	1.51	2.05	0.38
PC-Δy	1.01	0.93	0.90	0.60	1.06	1.13	1.27	1.31	0.40
PC-$y-\bar{y}^{1-Sided}$	1.07	1.07	1.11	0.89	1.02	1.02	1.14	1.21	0.34
PC-CapUtil	1.01	0.97	0.87	**0.54**	1.05	1.14	1.40	1.46	0.42
PC-ΔCapUtil	1.02	0.94	0.89	0.74	**0.99**	1.04	1.20	1.42	0.39
PC-Permits	1.21	1.08	1.28	1.03	1.01	1.11	1.18	1.31	0.32
PC-ΔPermits	1.14	0.95	1.11	1.06	1.00	1.00	1.00	0.90	0.32
PC-CFNAI	.	.	.	.	1.02	1.11	1.44	1.76	.

B. PCE Deflator (All Items)

	1970:I – 1983:IV				1984:I – 2004:IV				$\frac{RMSFE_{84-04}^{h=4}}{RMSFE_{70-83}^{h=4}}$
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	
AR(AIC) RMSFE	1.71	1.76	2.13	2.78	1.04	0.94	0.90	0.92	
Relative MSFEs									
AR(AIC)	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	1.00	0.42
AO	2.43	1.98	1.13	0.97	1.57	1.18	**0.74**	**0.79**	0.34
PC-u	0.78	0.92	0.99	0.69	**0.92**	1.04	1.26	1.77	0.47
PC-Δu	**0.74**	**0.83**	0.87	0.68	0.95	1.09	1.28	1.95	0.51
PC-$u-\bar{u}^{1-Sided}$	0.78	0.89	0.97	0.86	0.96	1.12	1.40	2.29	0.50
PC-Δy	0.84	0.85	0.84	**0.60**	1.05	1.09	1.19	1.46	0.50
PC-$y-\bar{y}^{1-Sided}$	0.83	0.94	0.91	0.76	1.05	1.11	1.26	1.65	0.50
PC-CapUtil	0.79	0.91	0.88	0.65	0.99	1.02	1.19	1.49	0.49
PC-ΔCapUtil	0.83	0.79	**0.82**	0.71	1.02	1.06	1.18	1.46	0.50
PC-Permits	1.01	0.91	0.96	0.81	1.02	1.06	0.94	1.20	0.42
PC-ΔPermits	1.14	1.04	1.00	0.83	1.02	1.00	1.01	1.08	0.42
PC-CFNAI	.	.	.	.	1.09	1.13	1.32	2.05	.

C. CPI

	1970:I – 1983:IV				1984:I – 2004:IV				$\frac{RMSFE_{84-04}^{h=4}}{RMSFE_{70-83}^{h=4}}$
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	
AR(AIC) RMSFE	2.17	2.24	2.67	3.66	1.41	1.29	1.27	1.24	
Relative MSFEs									
AR(AIC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.48
AO	2.54	1.95	1.11	0.95	1.78	1.31	**0.78**	**0.83**	0.40
PC-u	0.72	0.83	0.76	0.70	**0.98**	1.06	1.18	1.60	0.59
PC-Δu	0.66	**0.75**	**0.67**	0.61	1.02	1.08	1.25	1.56	0.65
PC-$u-\bar{u}^{1-Sided}$	0.68	0.80	0.75	0.87	1.02	1.10	1.33	1.91	0.64
PC-Δy	0.90	0.88	0.69	**0.58**	1.03	1.00	1.04	1.34	0.59
PC-$y-\bar{y}^{1-Sided}$	0.92	0.90	0.75	0.75	1.01	1.04	1.07	1.47	0.57
PC-CapUtil	**0.65**	0.82	0.73	0.64	1.05	**0.98**	1.02	1.20	0.56
PC-ΔCapUtil	0.71	0.82	0.68	0.62	1.09	1.01	1.05	1.28	0.59
PC-Permits	0.88	0.83	0.70	0.70	1.05	1.09	1.18	1.33	0.62
PC-ΔPermits	1.02	1.01	0.80	0.72	1.03	1.11	1.28	1.27	0.60
PC-CFNAI	.	.	.	.	1.07	1.11	1.30	1.86	.

Notes: See the notes to Table 1.

Table A.2: Autocorrelations of $\Delta\pi_t$ for the PCE deflator and CPI=U

	PCE (Core)		PCE (All)		CPI-U	
Lag	**1960:I – 1983:IV**	**1984:I – 2004:IV**	**1960:I – 1983:IV**	**1984:I – 2004:IV**	**1960:I – 1983:IV**	**1984:I – 2004:IV**
1	**-0.263** (0.102)	**-0.396** (0.109)	**-0.220** (0.102)	**-0.382** (0.109)	-0.202 (0.102)	-0.367 (0.109)
2	0.094 (0.109)	-0.143 (0.125)	-0.090 (0.107)	-0.174 (0.124)	-0.294 (0.106)	-0.183 (0.123)
3	-0.087 (0.110)	0.064 (0.127)	0.167 (0.108)	0.143 (0.127)	0.338 (0.114)	0.222 (0.126)
4	0.022 (0.110)	0.047 (0.127)	0.048 (0.110)	-0.023 (0.129)	0.014 (0.124)	-0.190 (0.131)
5	0.064 (0.110)	-0.082 (0.128)	-0.130 (0.111)	-0.115 (0.129)	-0.149 (0.124)	-0.018 (0.134)
6	-0.101 (0.111)	0.009 (0.128)	0.050 (0.112)	0.021 (0.130)	0.141 (0.126)	0.008 (0.134)
7	-0.026 (0.112)	0.165 (0.128)	0.024 (0.112)	0.242 (0.130)	0.028 (0.128)	0.178 (0.134)
8	-0.127 (0.112)	-0.158 (0.130)	-0.212 (0.112)	-0.131 (0.135)	-0.356 (0.128)	-0.065 (0.137)
$\hat{\sigma}_{\Delta\pi}$	1.10	0.73	1.32	1.15	1.85	1.55

Notes: See the notes to Table 2.

Table A.3. Pseudo Out-of-Sample Forecasting Performance of Additional Univariate Models: MSFEs, Relative to AR(AIC), PCE Deflator and CPI-U

(a) PCE-Core

	1970:I – 1983:IV				1984:I – 2004:IV			
Model	**h=1**	**h=2**	**h=4**	**h=8**	**h=1**	**h=2**	**h=4**	**h=8**
Recursive forecasts								
AR(AIC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AO	2.55	1.90	1.14	0.99	1.59	1.11	**0.71**	**0.59**
MA(1)	0.97	0.85	0.89	0.90	0.94	0.92	0.89	0.70
AR(4)	1.11	1.18	1.13	0.98	1.01	1.02	0.97	0.87
Rolling forecasts								
AR(AIC)	1.10	1.12	0.99	1.12	1.21	1.08	1.05	0.76
AR(4)	1.16	1.24	1.17	1.03	1.22	1.16	1.05	0.84
MA(1)	**0.96**	**0.83**	**0.87**	0.89	0.89	0.84	0.81	0.68
Nelson-Schwert								
NS77 MA(2)	1.08	1.01	0.99	**0.86**	**0.83**	**0.78**	0.75	0.65
Fixed-parameter comparisons								
UC-SV, $\gamma = 0.2$	0.93	0.84	0.88	0.88	0.91	0.84	0.78	0.67
MA(1) $\theta = 0.25$	0.92	0.82	0.85	0.86	0.96	0.95	0.92	0.73
MA(1) $\theta = 0.65$	1.16	1.05	1.01	0.88	0.81	0.75	0.71	0.61

(b) PCE-All

	1970:I – 1983:IV				1984:I – 2004:IV			
Model	**h=1**	**h=2**	**h=4**	**h=8**	**h=1**	**h=2**	**h=4**	**h=8**
Recursive forecasts								
AR(AIC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AO	2.43	1.98	1.13	0.97	1.57	1.18	0.74	0.79
MA(1)	0.87	0.89	0.84	**0.83**	0.98	0.94	0.90	0.87
AR(4)	0.99	1.09	1.05	1.00	0.99	0.99	1.00	0.94
Rolling forecasts								
AR(AIC)	1.03	0.96	0.96	0.97	1.04	1.00	0.94	0.96
AR(4)	1.01	1.13	1.08	1.08	1.04	1.01	0.98	0.97
MA(1)	**0.86**	**0.88**	**0.82**	**0.83**	0.93	0.86	0.80	0.83
Nelson-Schwert								
NS77 MA(2)	1.03	1.10	1.01	0.85	**0.87**	**0.79**	**0.73**	**0.73**
Fixed-parameter comparisons								
UC-SV, $\gamma = 0.2$	0.85	0.87	0.80	0.81	0.97	0.92	0.87	0.84
MA(1) $\theta = 0.25$	0.83	0.83	0.79	0.80	1.01	0.99	0.94	0.90
MA(1) $\theta = 0.65$	1.13	1.13	1.03	0.87	0.86	0.77	0.72	0.73

(c) CPI

	1970:I − 1983:IV				1984:I − 2004:IV			
Model	**h=1**	**h=2**	**h=4**	**h=8**	**h=1**	**h=2**	**h=4**	**h=8**
Recursive forecasts								
AR(AIC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AO	2.54	1.95	1.11	0.95	1.78	1.31	0.78	0.83
MA(1)	1.03	**1.03**	**0.87**	**0.83**	0.97	0.94	0.90	0.91
AR(4)	1.00	1.04	0.98	0.90	1.01	1.01	1.05	0.96
Rolling forecasts								
AR(AIC)	**0.92**	1.08	0.97	1.01	1.08	1.01	0.91	0.90
AR(4)	0.94	1.08	1.01	0.93	1.13	1.06	0.95	0.91
MA(1)	1.05	1.07	**0.87**	**0.83**	0.95	0.91	0.85	0.88
Nelson-Schwert								
NS77 MA(2)	1.12	1.18	1.04	0.86	**0.87**	**0.79**	**0.70**	**0.69**
Fixed-parameter comparisons								
UC-SV, $\gamma = 0.2$	1.03	1.07	0.86	0.83	1.08	1.09	1.06	1.04
MA(1) $\theta = 0.25$	0.98	0.97	0.82	0.81	1.02	1.01	0.96	0.97
MA(1) $\theta = 0.65$	1.24	1.19	1.05	0.88	0.88	0.79	0.71	0.72

Notes: See the notes to Table 4.

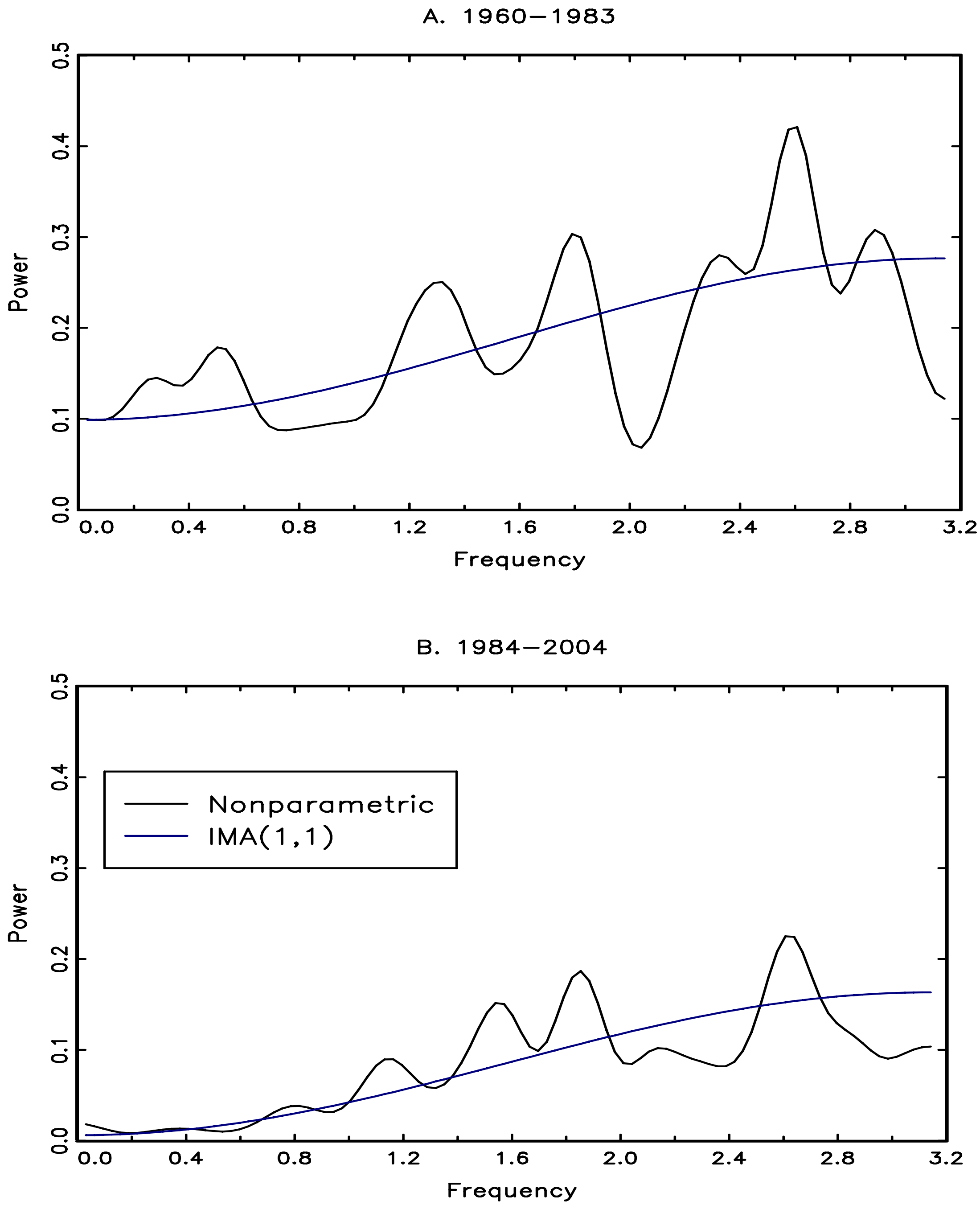
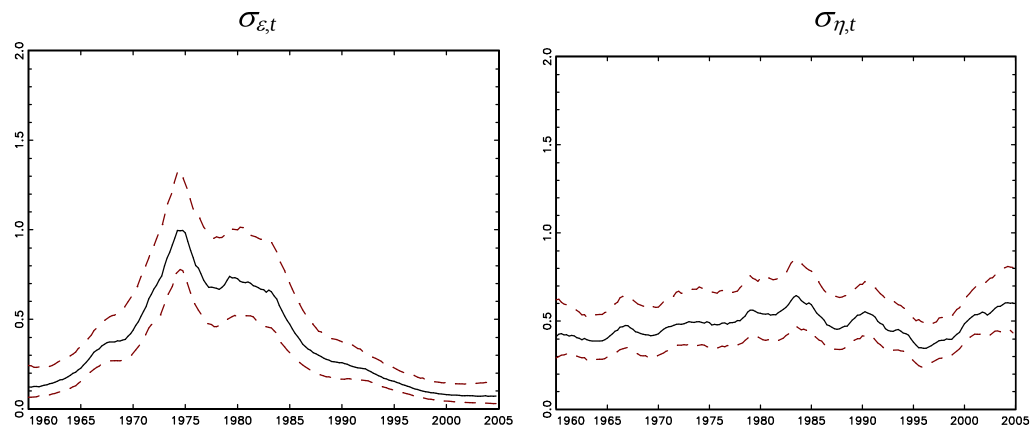
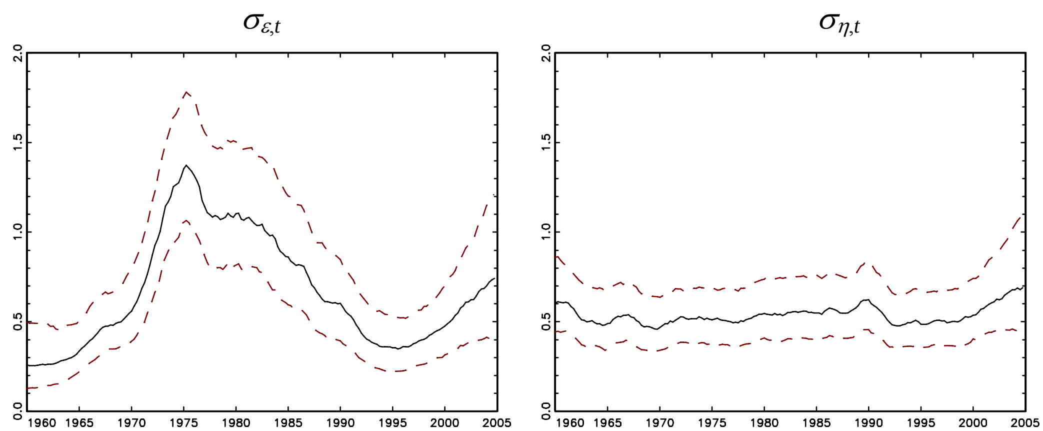


Figure A.1. Parametric and Nonparametric Estimates of the Spectrum of the first difference of PCE-core inflation, 1970-1983 and 1984-2004.

(a) PCE (Core)



(b) PCE (All Items)



(b) CPI-U

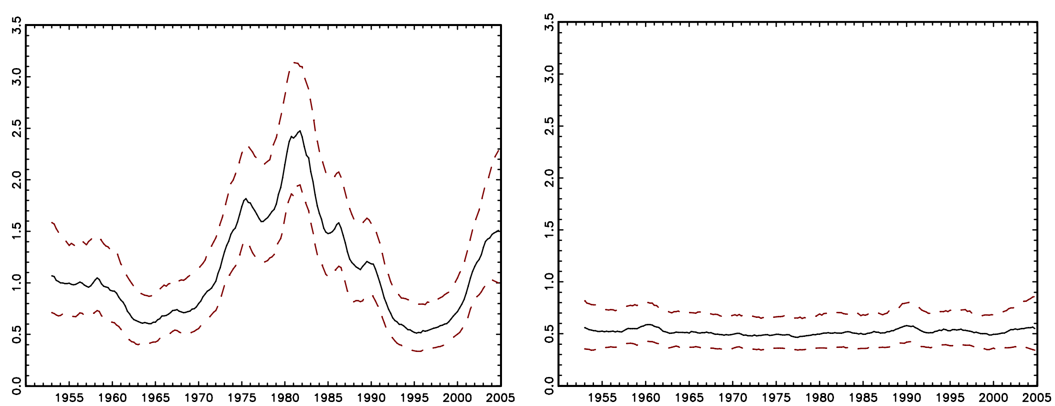


Figure A.2. Estimates of the standard deviations of the permanent and transitory innovations, and of the implied IMA(1,1) coefficient, using the TC-SV(.2) model: 16.5%, 50%, and 83.5% quantiles of the posterior distributions, 1959-2004.

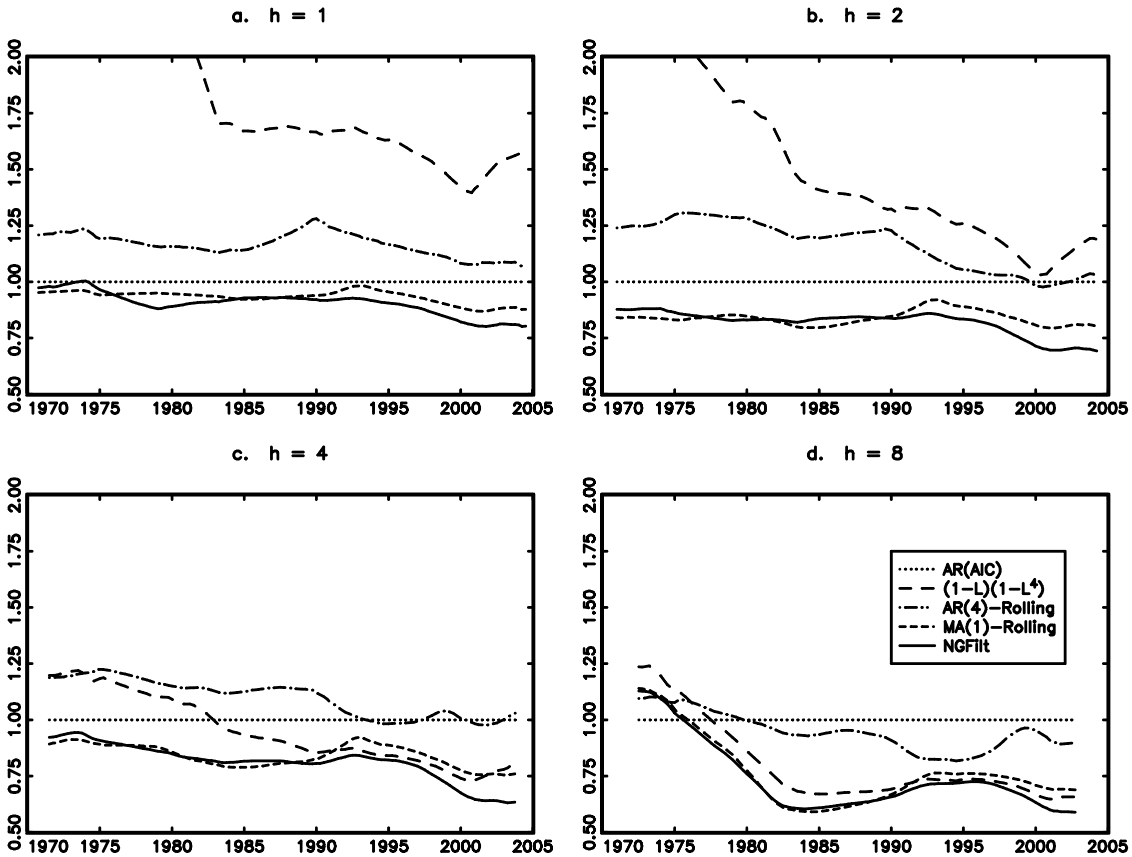


Figure A.3. Smoothed relative mean squared forecast errors of various forecasts, relative to the recursive AR(AIC) benchmark: PCE-core

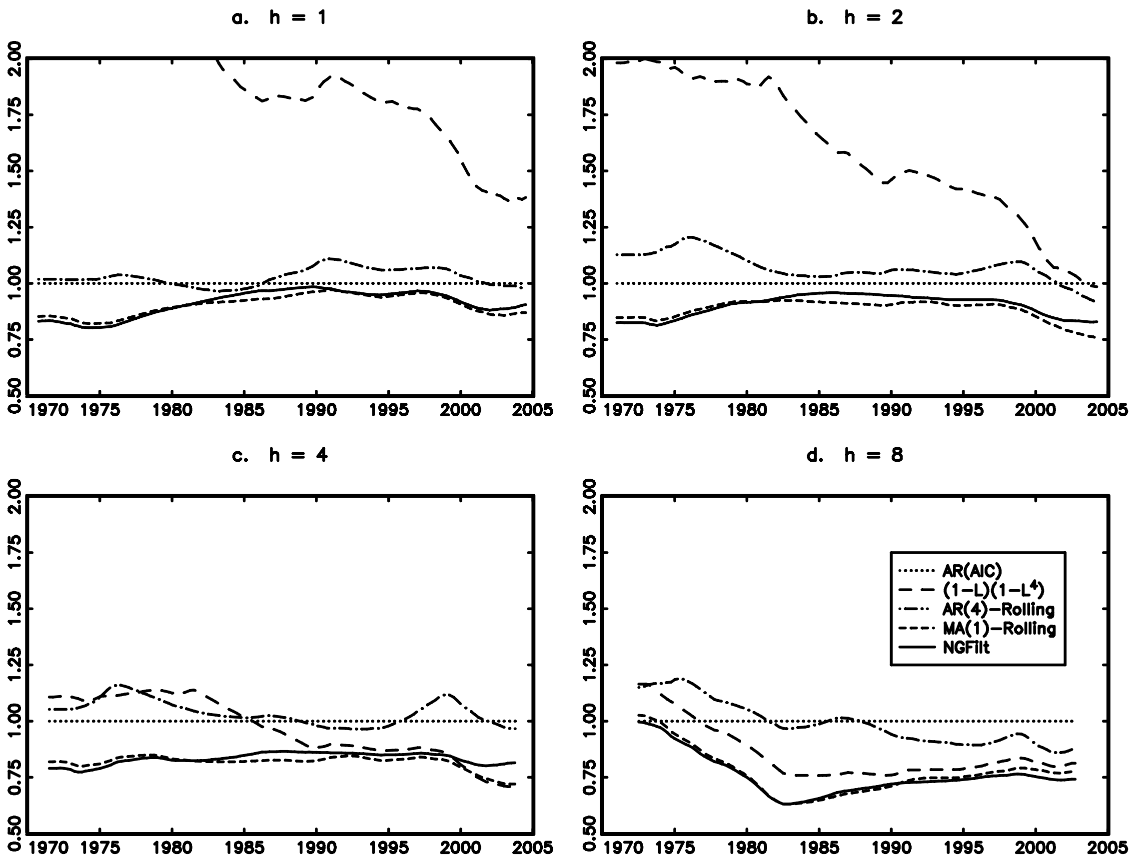


Figure A.4. Smoothed relative mean squared forecast errors of various forecasts, relative to the recursive AR(AIC) benchmark: PCE-all

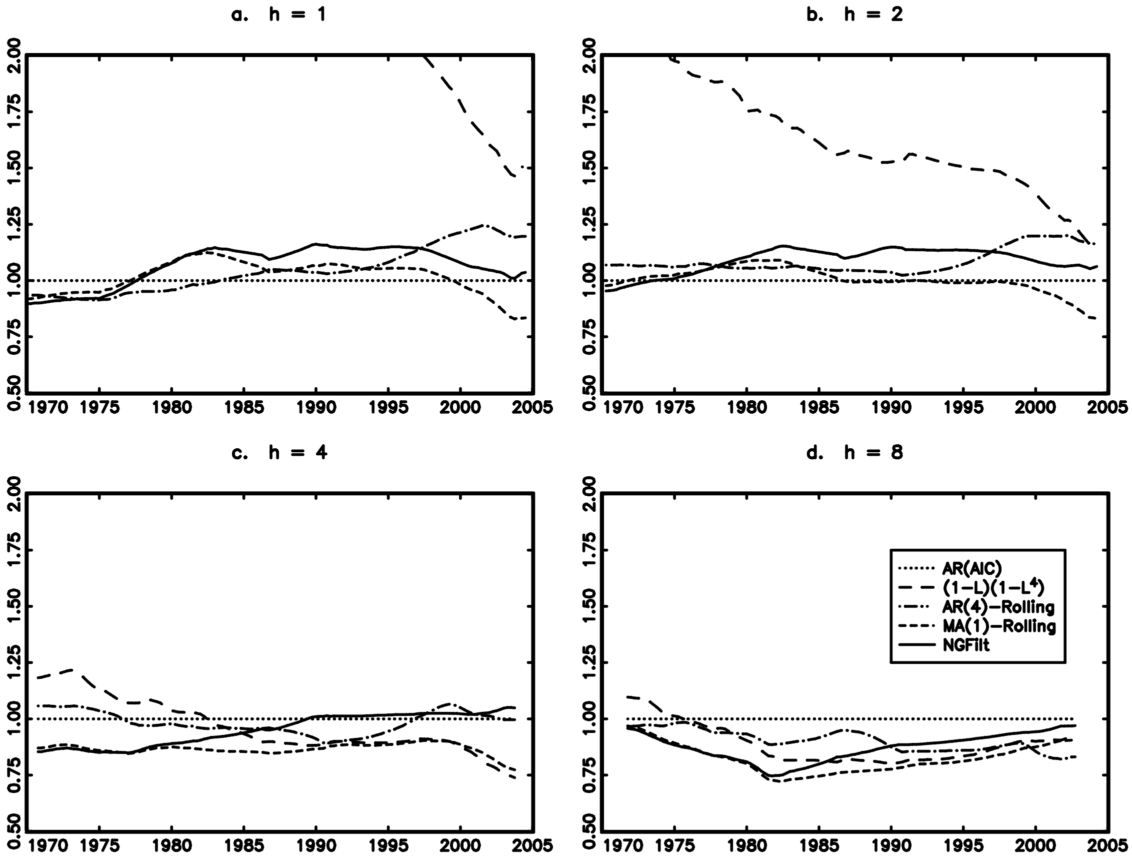


Figure A.5. Smoothed relative mean squared forecast errors of various forecasts, relative to the recursive AR(AIC) benchmark: CPI-U

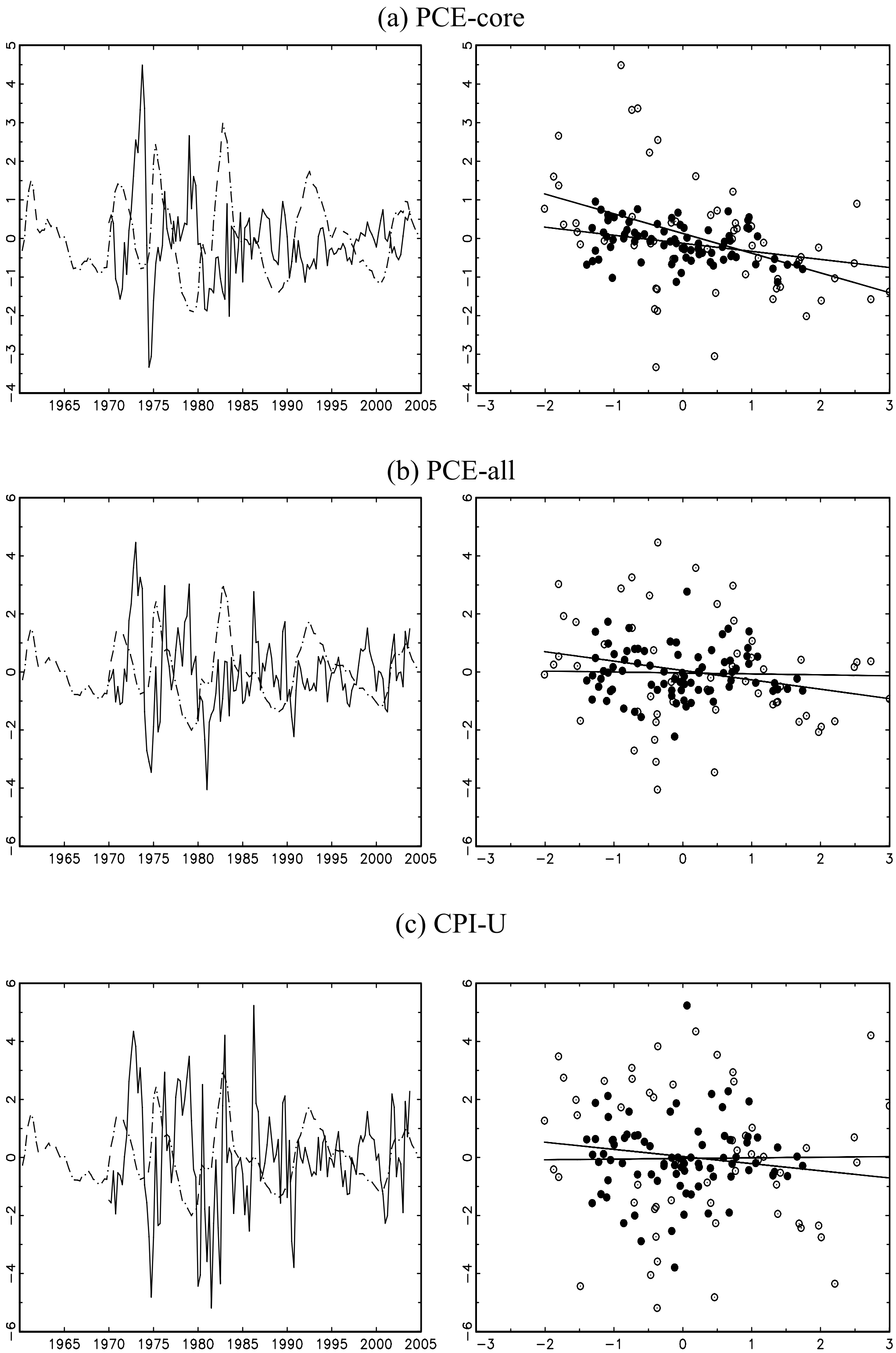


Figure A.6. Time series plot (left) and scatterplot of residual (right) from UC-SV model vs. the 1-sided unemployment gap. Scatterplot: open circles, 1970–83; filled circles, 1984–04.