The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time

Athanasios Orphanides*
Board of Governors of the Federal Reserve System

Simon van Norden
École des Hautes Études Commerciales, Montréal and CIRANO

September 2001

Abstract

A stable predictive relationship between inflation and the output gap, often referred to as a Phillips curve, provides the basis for empirical formulations of countercyclical monetary policy in many models. In this paper, we provide an empirical evaluation of the usefulness of alternative univariate and multivariate estimates of the output gap for predicting inflation. In-sample analysis based on ex post output gap measures suggests that many of the alternative estimates we examine appear to be quite useful for predicting inflation. However, examination of out-of-sample forecasts using real-time estimates of the same measures suggests that this predictive ability is mostly illusory. We find that inference based on ex post constructed output gap measures severely overstates their usefulness for predicting inflation in real time and, further, that real-time forecasts using the output gap are often less accurate than forecasts that abstract from the output gap concept altogether.

Keywords: Phillips curve, output gap, inflation forecasts, real-time data.

JEL Classification System: E37, C53.

Correspondence: Orphanides: Division of Monetary Affairs, Board of Governors of the Federal Reserve System, Washington, D.C. 20551, USA. Tel.: (202) 452-2654, e-mail: aorphanides@frb.gov. van Norden: H.E.C., 3000 Chemin de la Côte Sainte Catherine, Montréal QC, Canada H3T 2A7. Tel.: (514)340-6781, e-mail: simon.van-norden@hec.ca.

* This is a preliminary and incomplete draft prepared for the Federal Reserve Bank of Philadelphia Conference on Real Time Data Analysis. The opinions expressed are those of the authors and do not necessarily reflect views of the Board of Governors of the Federal Reserve System.

1 Introduction

A stable predictive relationship between inflation and a measure of deviations of aggregate demand from the economy's potential supply—the "output gap"—provides the basis for many formulations of activist countercyclical stabilization policy. Such a relationship, referred to as a Phillips curve, is often seen as a helpful guide for policymakers aiming to maintain low inflation and stable economic growth. According to this paradigm, when aggregate demand exceeds potential output, the economy is subject to inflationary pressures and inflation should be expected to rise. Under these circumstances, policymakers might wish to adopt policies restricting aggregate demand aiming to contain the acceleration in prices. Similarly, when aggregate demand falls short of potential supply, inflation should be expected to fall, prompting policymakers to consider adoption of expansionary policies to restore stability.

Regardless of the analytical usefulness or the theoretical validity of a presumed predictive relationship between a concept of the output gap and inflation, however, the *practical* usefulness of such a relationship is largely an empirical matter. Even under the presumption that a stable predictive relationship is present in the data, a number of issues may complicate its use in practice. The appropriate empirical definition of "potential output"—and the accompanying "output gap"—that might be useful in practice is far from clear. For any given empirical definition of the gap, the exact form of its empirical relationship with inflation cannot be known a priori and would need to be determined from the data. Further, even if we were to assume that the proper concept and empirical relationship are identified, the operational usefulness of this predictive relationship would be subject to the availability of reliable estimates of the relevant gap concept in real time, when the desired inflation forecasts are required. But as is well known, real-time estimates of the output gap are generally subject to significant revisions.[1] The subsequent evolution of the economy provides

[1] For example, see Orphanides and van Norden (2002).

useful information for determining which part of the business cycle the economy was in at a particular point in time—information which leads to improved estimates of the gap. As a result, considerable uncertainty regarding the value of the gap remains even long after it would be needed for forecasting inflation. This, in turn, raises questions regarding the empirical usefulness of the output gap for forecasting inflation in real time.

In this paper we assess the usefulness of alternative univariate and multivariate methods for estimation of the output gap for predicting inflation, paying particular attention to the distinction between *suggested usefulness*—based on in-sample historical analysis— and *operational usefulness*—based on simulated real-time out-of-sample analysis. First, using in-sample analysis based on ex post estimates of the output gap, we confirm that some appear to be useful for predicting inflation. This is as would be expected since the implicit Phillips curve relationships recovered in this manner are similar to the relationships commonly found in empirical macroeconometric models. However, the ability to explain inflation ex post does not imply an operational ability to forecast inflation. To assess the latter, we generate out-of-sample forecasts based on *real-time* output gap measures; those constructed using only data (and parameter estimates) available at the time forecasts are generated. For this exercise, we rely on the real-time dataset for macroeconomists which was created and is maintained by the Federal Reserve Bank of Philadelphia.[2]

Our findings based on this real-time analysis suggest that the predictive ability of output gap measures mostly illusory. Forecasts based on ex post estimates of the output gap severely overstate the gap's usefulness for predicting inflation. Further, real-time forecasts using the output gap are often less accurate than forecasts that abstract from the output gap concept altogether. These results bring into question the reliability and practical usefulness of inflation forecasts based on output gaps.

[2]See Croushore and Stark (forthcoming) for background information regarding this database.

2 Related Literature

(To be added. Related work on inflation forecasting and unemployment, e.g. Stock and Watson (1999), and relationship with NAIRU estimation, e.g. Staiger, Stock and Watson (1997a,b). Related work on unreliability of real-time output gap estimates.)

3 Trends and Cycles Ex Post and in Real Time

One way to define the output gap is as the difference between actual output and an underlying unobserved trend towards which output would tend to revert in the absence of business cycle fluctuations. Let q_t denote the (natural logarithm of) actual output during quarter t, and μ_t its trend. Then, the output gap, y_t can be defined as the cycle component resulting from the decomposition of output into a trend and cycle component:

$$q_t = \mu_t + y_t$$

Since the underlying trend is unobserved, its measurement, and the resulting measurement of the output gap, very much depends on the choice of estimation method, underlying assumptions and available data that are brought to bear on the measurement problem. For any given method, simple changes in historical data and the availability of additional data can change, sometimes drastically, the resulting estimates of the cycle for a given quarter. As a result, examination and interpretation of statistical relationships between the "output gap" and other variables, such as inflation, requires additional specificity regarding the temporal perspective from which the relationship is examined.

To illustrate this issue figures 1 and 2 provide some comparisons of output gap measures obtained using the Hodrick-Prescott (HP) filter using alternative information sets.[3] The solid line in the top panel of figure 1 denotes the output gap obtained with our "final" dataset

[3] We selected the HP filter (Hodrick and Prescott, 1997) for this illustration because of its popularity and simplicity which have made it a focus of extensive analysis and a benchmark for comparisons with alternative detrending techniques, both univariate and multivariate. See, for example, Harvey and Jaeger (1993), King and Rebelo (1993), Cogley and Nason (1995), Kozicki (1999), and Christiano and Fitzgerald (1999).

with data ending in 1999Q4 as published in 2000Q1. The dotted line, instead, shows real-time estimates of the gap, as could be estimated with the historical data available at the time data first became available for that quarter. Thus, the real-time estimate for 1969Q1 was obtained by applying the Hodrick-Prescott filter to the data available in 1969Q2, when output figures for 1969Q1 were first released. Similarly, the real-time estimate for 1995Q4 was obtained by applying the Hodrick-Prescott filter to the data available in 1996Q1. The bottom panel provides a similar comparison of the four-quarter moving average of the output gap, as estimated over history and in real-time. Comparison of the series in either panel indicates that the resulting real-time and final series for the output gap exhibit significant differences. The series roughly agree on the timing of periods when output was significantly above or below its trend—as defined by the filter. But, as is also apparent from the figure, the real-time and final series frequently do not even agree on whether the output gap is positive or negative.

Figure 2 illustrates this difficulty in greater detail for two specific episodes. The top panel compares the historical estimates of the output gap as could be constructed in 1969Q1 with the final estimates. As can be seen, the real-time estimates as could have been constructed at the beginning of 1969 based on this method would have suggested that the economy was operating below its trend for the previous two years. But based on the ex post estimates, the output gap during the previous year was positive. The implications of this difference for a forecasting exercise are quite clear. Presuming the presence of a positive predictive relationship between the output gap and inflation, the ex post estimates would have suggested inflationary pressures. But the real-time estimates would have suggested the opposite, instead. The bottom panel provides a similar comparison where the opposite conflict is apparent. Historical estimates of the output gap as could be constructed in 1996Q1 would have indicated an overheated economy during the previous year, whereas the final estimates suggest output was below its trend, instead.

Further evidence of the difference between historical and real-time estimates of output gaps has been presented by Orphanides and van Norden (1999, 2002). In Table 1, we present some of the summary reliability indicators they examine for twelve alternative measures of the output gap, which we employ in our analysis. (These are described in greater detail below.) These results show that revisions in real-time estimates are often of the same magnitude as the historical estimates themselves and confirm that historical and real-time estimates frequently have opposite signs for many of the alternative methods.

As these examples illustrate, the presence of a predictive relationship between the output gap and inflation based on ex post estimated output gap measures might not be sufficient to assess whether the output gap could provide useful information for forecasting inflation in real time. Importantly, this is a difficulty that would apply even if such a predictive relationship were precisely estimated and known to be quite useful in-sample. Of course, if this relationship were not known exactly, its estimation—which would also need to be performed in real time—would present additional some difficulties. Econometric estimates would obviously also change with the evolving renditions of historical output gaps, even for a relationship estimated over a fixed sample.

4 A Forecasting Experiment

The results above suggest that ex post estimates of output gaps at a point in time may differ substantially from estimates which could be made without the benefit of hindsight. We now turn to consider what effect, if any, such differences could have on their ability to forecast inflation. The remainder of this section discusses the methodology used to investigate this question. We begin by describing the data sources used, and we then discuss the measurement of the output gap in more detail. Thereafter, we detail how the forecasting power of these output gap estimates is gauged.

4.1 Data Sources and Vintages

We use the term *vintage* to describe the values for data series as published at a particular point in time. Most of our data is taken from the real-time data set compiled by Croushore and Stark (1999); we use the quarterly vintages from 1965Q1 to 1999Q4 for real output. Construction of the output series and its revision over time is further described in Orphanides and van Norden (1999, 2002).

We use 2000Q1 data as “final data” recognizing, of course, that “final” is very much an ephemeral concept in the measurement of output.

To measure inflation, we use the quarterly rate of inflation in the consumer price index (CPI). We use this both for our forecasting experiments and also to estimate measures of the output gap based on multivariate models that include inflation. For all of our analysis, we rely on the consumer price index (CPI) as available in 2000Q1. CPI data do not generally undergo a similar revision process as the output data. The major source of revision is changes in seasonal factors most noticeable at a monthly frequency. We therefore use the 2000Q1 vintage of CPI data for all the analysis which allows us to focus our attention on the effects of revisions in the output data and the estimation of the output gap in our analysis. One of our models (Blanchard-Quah) also uses data on interest rates which do not undergo any revisions at all.

4.2 Measuring Output Gaps

We construct ensembles of output gaps estimates of varying vintages. Each output gap vintage uses precisely one vintage of the output data. An estimated output gap is called a final estimate if it uses the final data vintage.

These ensembles of varying vintages of output gap estimates were constructed for each of a number of different output gap estimation techniques The alternative methods are detailed in the Appendix. Some, such as the linear or the quadratic trend, are based on purely deterministic detrending methods. Some, such as the Hodrick-Prescott filter, do not directly rely on statistical model-fitting. Five are estimated unobserved components, of which three (Watson, Harvey-Clark and Harvey-Jaeger) are univariate models and two (Kuttner and Gerlach-Smets) are bivariate models, using data on both output and prices. The remaining models are all univariate with the exception of the Blanchard-Quah method, which uses a trivariate structural VAR.

Note that all the output gap estimation techniques (aside from the Hodrick-Prescott filter) require that one or more parameters be estimated to fit the data. Such estimation was repeated for every combination of technique and vintage. This means, for example, that in constructing output gap vintages from an unobserved components model spanning the thirty year period 1969Q1-1998Q4 (120 quarters), we reestimate the model's parameters 120 times, and then store 120 series of filtered estimates.

4.3 Forecasting Specification

We restrict attention to linear specifications. Let $\pi_t^h = (400/h)(log(P_t) - log(P_{t-h}))$ denote inflation over h quarters ending in quarter t, at an annual rate. (The quarterly rate of inflation is simply $\pi_t \equiv \pi_t^1 = (400)(log(P_t) - log(P_{t-1}))$.) We are mainly interested in examining forecasts of inflation over a one year horizon. Thus, given data for quarter t and earlier periods, our objective is to forecast $\pi_{t+4}^4 = (\pi_{t+4} + \pi_{t+3} + \pi_{t+2} + \pi_{t+1})/4$. We note that because of reporting lags, information for quarter t is not available before quarter $t+1$. Thus, a four-quarter ahead forecast is a forecast five quarters ahead of the last quarter for which actual data are available. The forecasting relationship we examine is thus:

$$\pi_{t+4}^4 = \alpha + \sum_{i=0}^{n} \beta_i \pi_{t-i} + \sum_{i=0}^{m} \gamma_i y_{t-i} + e_{t+4} \tag{1}$$

where n and m denote the number of lags of inflation and the output gap in the equation.

Given a concept of the output gap, two issues complicate the interpretation of how we could obtain inflation forecasts using equation (1). First, since the most suitable number of lags of inflation and the output gap n and m, and the parameters of the equation are not known a priori, these need to be estimated with available data. As our sample increases and additional data become available we would expect, of course, that these estimates would change.[4] Second, the estimates of the historical output gap available up to some specific period are revised and also change over time. This in turn, has two possible effects. First, this alone can influence the determination of the most suitable number of lags and the estimated parameters of equation (1)—for any fixed estimation sample. Second, given some fixed values of the parameters of the equation, the implied forecasts corresponding to the revised estimates of the output gap would be different as well.

In examining the usefulness of the output gap for predicting inflation using equation (1), we thus perform two different experiments for every output gap estimation technique we examine. First, we examine the in-sample fit of the the data, using final estimates of the output gap to both estimate (1) and compute its fitted values. Second, we simulate a real-time out-of-sample forecasting exercise. In this case, in each quarter, t we use the tth vintage of the output gap series to estimate (1) (which includes determining its lag lengths m and n) and to generate its implied forecast.

To provide a benchmark for comparison, we estimate a univariate forecasting model of inflation based on equation (1) but omitting the output gaps. Again, we do this twice, first in-sample and second in simulated real-time, re-estimating the model with each additional observation.

This experiment is designed to mimic in a simple way the problem facing a policymaker who wishes to forecast inflation in real time. Of course, the forecasting problem faced by policymakers in practice is more complex than the one we consider. One obvious and important difference is that the information set available to policymakers is much richer. It is therefore possible that output gaps might improve on simple univariate forecasts of

[4] The results we report in subsequent sections use the Ng-Perron approach of determining lag length within a general-to-specific testing framework and using t-ratio tests to determine the last significant lag. A full sensitivity analysis will be done to determine the impact of using alternative lag selection rules such as AIC or BIC. However, preliminary limited experiments seemed to suggest that such changes would not have a major impact on the results.

inflation but not on forecasts using a broader range of inputs. For this reason, we feel that the experiment we perform is only a weak test of the utility of empirical output gap models.

In addition, we examine two sets of forecasts obtained using equation (1) but replacing the output gaps with either real or nominal output growth not subjected to any prior filtering and/or smoothing. As van Norden (1995) explains, using output growth in this way can be interpreted as implicitly defining an estimated output gap as a one-sided filter of output growth with weights based on the estimated coefficients of equation (1).[5] On the other hand, this approach does not rely nor require prior estimation of an output gap measure per se and is therefore simpler. The resulting forecasts should give the best linear unbiased predictors of future inflation (since OLS is BLUP). Since our output gap measures use the same information set (past prices and the current vintage of output) as these unrestricted forecasting equations, comparing their forecast performance aids in isolating the usefulness of the economic structure (or other restrictions) embedded in our output gap measures.

5 Baseline Results

5.1 In-sample

To examine the in-sample performance of alternative measures of the gap, we estimated equation (1) using observations from 1955:1 to 1998:3. To allow for direct comparison with the simulated real-time forecasts, we examine its fit only over the period starting in 1969:1—the first quarter for which we also have a simulated real-time forecast. Table 2 presents the root-mean-square errors (RMSE) of the resulting in-sample forecasts of inflation. The first 12 rows in the table reflect alternative detrending methods, and the last three show statistics for the autoregressive forecast benchmark and the forecasts based on current and lagged real or nominal income growth instead of a pre-defined output gap measure. For each method, the first column shows the resulting forecast RMSE for the whole evaluation period, 1969-1998 and the remaining two columns show the same statistic for two subsamples, 1969-1983

[5] van Norden refers to these estimates as TOFU (Trivial Optimal Filter - Unrestricted).

and 1984-1998. The break follows the one examined by Stock and Watson (1999) in their study of inflation forecasts and splits the evaluation sample in two parts with roughly equal observations. The two subsamples also correspond, respectively, to a period of relatively high and relatively low variability in inflation. As can be seen, the autoregressive forecast (AR) has a RMSE of about 1.9 percent for the whole period but much higher (2.3 percent) during the first part of the sample and much lower (1.3 percent) during the second half. A quick comparison of the RMSE of alternative methods with the AR in column 1 indicates that *all* 12 of the alternative detrending techniques suggest improved forecast errors for the complete sample and 10 of the 12 also suggest improvements for both subsamples. Use of real or nominal growth also appears to improve the inflation forecasts relative to the autoregression.

To assess whether these improvements are statistically significant, we computed a modified Diebold-Mariano statistic for the alternative forecasts. Table 3 shows the ratio of the RMSEs shown in Table 2 to the RMSE of the AR forecasts, and notes when the modified Diebold-Mariano statistic indicated that the null hypothesis that the RMSEs are the same was rejected at the 5% or 10% percent level. As can be seen, according to this test, 10 of the 12 detrending methods indicate that the improvement in forecasts for the whole sample appears significant at the 5% percent level and only one (the Harvey-Jaeger technique) failed to reject the hypothesis of no improvement at the 10% percent level. An interesting aspect of the evaluation for the two sub-samples, however, indicates that most of the improvement appears to be concentrated in the first half of the sample, when inflation was more volatile. As judged by the modified Diebold-Mariano statistic, the improvement in forecasts is statistically significant at the 5% or 10% level in both subsamples for only five of the twelve methods. The overall forecast improvement associated with using several of the output gap measures examined in Table 3, however, provides evidence of the potential usefulness of these output gap measures for forecasting inflation.

5.2 Real-Time

Next, we ran the simulated out-of-sample forecasting experiment. In each quarter t starting with 1969:1, we re-estimated equation (1) with data vintage t starting from 1955:1.[6] We then used equation (1) to obtain the inflation forecast corresponding to that method for that quarter. We repeated the procedure for each quarter up to 1998:3 and for each method.

The results are presented in Tables 4 and 5. These are directly comparable to tables 2 and 3, respectively. Comparison of the entries in Table 4 with respective entries in Table 2 indicates that the forecast performance of all methods appears markedly worse in real-time than in-sample. For the AR forecast benchmark, for example, the RMSE in real-time for the 1969-1998 evaluation sample is 2.3 percent, compared to the in-sample value of 1.9 percent. The forecast deterioration when we compare in-sample and real-time results, however, appears more severe for output gap based forecasts. Looking at the ratios of the RMSE relative to the AR forecast shown in Table 5 for the 1969-1998 period, we note that six of the twelve output gap methods indicate that the output-gap-based forecasts are worse, on average, than the autoregressive forecasts. Of the six that indicate some improvement, none indicates that this improvement is statistically significant even at the 10% level, based on the modified Diebold-Marianno statistic. Interestingly, forecasts based on real and nominal output growth appear to deteriorate somewhat less than those based on output gaps and only forecasts based on nominal output growth appear to significantly improve on the autoregressive forecasts for the full sample. However, examining subsamples, we note that this improvement is only evident in the first half of the sample and is not apparent in the second.

The results in Table 5 suggest that few output gap measures are of practical use, at the margin, for improving real-time forecasts of inflation. None are significantly useful. None

[6] Our output series start in 1947Q1, so the choice of 1955Q1 is somewhat arbitrary. Preliminary experiments shows that the results were somewhat sensitive to this choice. For example, starting earlier in the 1950s, resulted in inflation foreacting models that suggested worse fits and worse out-of-sample forecasting performance, overall. We plan to report sensitivity results in subsequent versions of the paper.

appear to improve the forecasts very much. This despite the fact that the benchmark (AR) forecast is trivially simple and uses unrealistically little information.

It is of interest to also examine the forecasting performance of the output gaps relative to the forecasts using real or nominal growth directly. Table 6 compares the performance of alternative methods for the whole sample for the three alternatives. The first column is replicated from the first column of table 5, using the AR forecast as a benchmark. The second and third column use, instead, the real and nominal growth based forecasts as benchmarks. As can be seen, using the real or nominal growth-based forecasts as benchmarks suggests even more disappointing results regarding the usefulness of output gaps for predicting inflation. Only one of the twelve output gap measures suggests any overall improvement in the forecasts over those based on real growth and none suggests any improvement over forecasts based on nominal growth.

6 Conclusion

Forecasting inflation is a difficult but essential task for the successful implementation of monetary policy. The hypothesis that a stable predictive relationship between inflation and the output gap—a Phillips curve—is present in the data, suggests that output gap measures could be useful for forecasting inflation. This has served as the basis for empirical formulations of countercyclical monetary policy. We find that many alternative measures of the output gap *appear* to be quite useful for forecasting inflation, on the basis of in-sample analysis. That is, a historical Phillips curve is suggested by the data, and ex post estimates of the output gap are useful for understanding historical movements in inflation. However, this suggested usefulness does not imply a similar operational usefulness. Our simulated real-time forecasting experiment suggests, instead, that this predictive ability is mostly illusory. These results bring into question the practical usefulness of output-gap-based Phillips curves for forecasting inflation and the monetary policy process.

References (Incomplete)

Baxter, Marianne; King, Robert G., "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series" National Bureau of Economic Research Working Paper: 5022, 1995.

Beveridge, S and C. R. Nelson, "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics,* 7, 151-174, 1981.

Cayen, Jean-Philippe "Fiabilité des estimations de l' écart de production au canada.? École des Hautes Études Commerciales (Montréal), memoire, 2001.

Christiano, Lawrence J., and Terry J. Fitzgerald, "The Band Pass Filter," NBER Working Paper, No. 7257, July 1999.

Clark, Peter K., "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics* 102(4), 1987, 797-814.

Cogley, T. and J. Nason, "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control*, 19(1-2), 1995, 253-78.

Croushore, Dean and Tom Stark, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, (forthcoming).

Diebold, Francis X. and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 1995, 253-265.

Gerlach, Stefan and Frank Smets, "Output Gaps and Inflation: Unobserable-Components Estimates for the G-7 Countries." Bank for International Settlements mimeo, Basel 1997.

Harvey, Andrew C., "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3 (1985). 216-227.

Harvey, A. C. and A. Jaeger, "Detrending, Stylized Facts, and the Business Cycle," *Journal of Applied Econometrics*, 8 (1993), 231-247.

Hodrick, R, and E. Prescott, "Post-war Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking,* 29, 1997, 1-16.

King, Robert G. and Sergio Rebelo, "Low Frequency Filtering and Real Business Cycles." *Journal of Economic Dynamics and Control*, 17(1-2), 1993, 207-31.

Kozicki, Sharon, "Multivariate Detrending Under Common Trend Restrictions: Implications for Business Cycle Research, *Journal of Economic Dynamics and Control*, 23(7)

June 1999, 997-1028.

Kuttner, Kenneth N., "Estimating Potential Output as a Latent Variable," *Journal of Business and Economic Statistics*, 12(3), 1994, 361-68.

Orphanides, Athanasios and Simon van Norden, "The Reliability of Output Gap Estimates in Real Time," Finance and Economics Discussion Series 1999-38, August 1999.

Orphanides, Athanasios and Simon van Norden, "The Unreliability of Output Gap Estimates in Real Time," *Review of Economics and Statistics*, forthcoming 2002.

Runstler, Gerhard, "Are Real-Time Estimates of the Output Gap Reliable? An Application to the Euro Area," European Central Bank mimeo, 2001.

St-Amant, Pierre and Simon van Norden, "Measurement of the Output Gap: A discussion of recent research at the Bank of Canada," Bank of Canada Technical Report No. 79, 1998.

Staiger, Douglas, James H. Stock, and Mark W. Watson, "How Precise are Estimates of the Natural Rate of Unemployment?" in Romer, Christina and David Romer, eds. *Reducing Inflation: Motivation and Strategy*, Chicago: University of Chicago Press, 1997a.

Staiger, Douglas, James H. Stock, and Mark W. Watson, "The NAIRU, Unemployment and Monetary Policy," *Journal of Economic Perspectives* 11(1), Winter 1997, 33-49.

Stock, James H. and Mark W. Watson, "Forecasting Inflation," *Journal of Monetary Economics,* 44, 293-335, 1999.

van Norden, Simon, "Why is it so hard to measure the current output gap?" Bank of Canada mimeo, 1995.

Watson, Mark W., "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 1986, 49-75.

Appendix: Alternative Detrending Methods

A detrending method decomposes the log of real output, q_t, into a trend component, μ_t, and a cycle component, z_t.

$$q_t = \mu_t + z_t \tag{A.1}$$

Some methods use the data to estimate the trend, μ_t, and define the cyclical component as the residual. Others specify a dynamic structure for both the trend and cycle components and estimate them jointly. We examine detrending methods that fall into both categories.

A.1 Deterministic Trends

The first set of detrending methods we consider assume that the trend in (the logarithm of) output is well approximated as a simple deterministic function of time. The linear trend is the oldest and simplest of these models. The quadratic trend is a popular alternative.

Because of the noticeable downturn in GDP growth after 1973, another simple deterministic technique is a breaking linear trend that allows for the slowdown in that year. Our implementation of the breaking trend method incorporates the assumption that the location of the break is fixed and known. Specifically we assume that a break in the trend at the end of 1973 would have been incorporated in real time from 1977 on. As discussed in Orphanides and van Norden (1999) this conforms with the debate regarding the productivity slowdown during the 1970s.

A.2 Unobserved Component Models and the Hodrick–Prescott Filter

Unobserved component (UC) models offer a general framework for decomposing output into an unobserved trend and a cycle, allowing for an assumed dynamic structure for these components.

This framework can also nest smoothing splines, such the popular filter proposed by Hodrick and Prescott (1997) (the HP filter). We implement the HP filter, following Harvey and Jaeger (1993) and King and Rebelo (1993), by writing it in its unobserved components form. Assuming that the trend in (1) follows:

$$(1 - L)^2 \mu_t = \eta_t \tag{A.2}$$

the HP filter is obtained from (A.1) and (A.2) under the assumption that z_t and η_t are mutually uncorrelated white noise processes with a fixed relative variance q. We set q to correspond to the standard application of the HP filter with a smoothing parameter of 1600.

UC models also permit more complex dynamics to be estimated, and we examine two such alternatives, by Watson (1986) and by Harvey (1985) and Clark (1987). The Watson model modifies the linear level model to allow for greater business cycle persistence. Specifically, it models the trend as a random walk with drift and the cycle as an AR(2) process:

$$\mu_t = \delta + \mu_{t-1} + \eta_t \tag{A.3}$$

$$z_t = \rho_1 \cdot z_{t-1} + \rho_2 \cdot z_{t-2} + \varepsilon_t \tag{A.4}$$

Here ε_t and η_t are assumed to be i.i.d mean-zero Gaussian and mutually uncorrelated and δ, ρ_1 and ρ_2, and the variances of the two shocks are parameters to be estimated (5 in total).

The Harvey-Clark model similarly modifies the local linear trend model:

$$\mu_t = g_{t-1} + \mu_{t-1} + \eta_t \tag{A.5}$$

$$g_t = g_{t-1} + \nu_t \tag{A.6}$$

$$z_t = \rho_1 \cdot z_{t-1} + \rho_2 \cdot z_{t-2} + \varepsilon_t \tag{A.7}$$

Here η_t, ν_t, and ε_t are assumed to be i.i.d mean-zero Gaussian and mutually uncorrelated processes and ρ_1 and ρ_2 and the variances of the three shocks are parameters to be estimated (5 in total).

A.3 Unobserved Component Models with a Phillips Curve

Multivariate formulations of UC models attempt to refine estimates of the output gap by incorporating information from other variables linked to the gap. We consider two models which add a Phillips curve to the univariate formulations described above; those of Kuttner (1994) and Gerlach and Smets (1997).

Let π_t be the quarterly rate of inflation. The Kuttner model adds the following Phillips curve equation to the Watson model:

$$\Delta\pi_t = \xi_1 + \xi_2 \cdot \Delta q_t + \xi_3 \cdot z_{t-1} + e_t + \xi_4 \cdot e_{t-1} + \xi_5 \cdot e_{t-2} + \xi_6 \cdot e_{t-3} \tag{A.8}$$

The Gerlach-Smets model modifies the Harvey-Clark model by adding the similar Phillips curve:

$$\Delta\pi_t = \phi_1 + \phi_2 \cdot z_t + e_t + \phi_3 \cdot e_{t-1} + \phi_4 \cdot e_{t-2} + \phi_5 \cdot e_{t-3} \tag{A.9}$$

In each case the shock e_t is assumed i.i.d. mean zero and Gaussian. In the Gerlach-Smets model, e_t is also assumed uncorrelated with shocks driving the dynamics of the trend and cycle components of output in the model. Thus, by adding the Phillips curve, the Gerlach-Smets model introduces an additional six parameters that require estimation ($\{\phi_1, ..., \phi_5\}$ and the variance of e_t). The Kuttner model also allows for a non-zero correlation between e_t and the shock to the cycle, η_t. Thus, it introduces eight additional parameters that require estimation ($\{\xi_1, ..., \xi_6\}$, the variance of e_t and its covariance with η_t.)

A.4 The Band-Pass Filter

Another approach to cycle-trend decomposition is via the use of band-pass filters in the frequency domain. The clearest exponent of this approach is Baxter and King (1999), who suggest the use of truncated versions of the ideal (and therefore infinitely long) filter with a band passing fluctuations with durations between 6 and 32 quarters in length. Stock and Watson (1998) adapt this for use at the end of data samples by padding the available observations with forecasts from a low-order AR model fit to the data series. Following Stock and Watson, we use a filter 25 observations in length and pad using an AR(4) forecast.

A.5 The Beveridge-Nelson Decomposition

Beveridge and Nelson (1981) consider the case of an ARIMA(p,1,q) series, y, which is to be decomposed into a trend and a cyclical component. For simplicity, we can assume that all deterministic components belong to the trend component and have already been removed from the series. Since the first-difference of the series is stationary, it has an infinite-order MA representation of the form

$$\Delta y_t = \varepsilon_t + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \cdots = e_t \tag{A.10}$$

where ε is assumed to be an innovations sequence. The change in the series over the next s periods is simply

$$y_{t+s} - y_t = \sum_{j=1}^{s} \Delta y_{t+j} = \sum_{j=1}^{s} e_{t+j} \tag{A.11}$$

The trend is defined to be

$$\lim_{s\to\infty} E_t(y_{t+s}) = y_t + \lim_{s\to\infty} E_t(\sum_{j=1}^{s} e_{t+j}) \tag{A.12}$$

From equation 6, we can see that

$$E_t(e_{t+j}) = E_t(\varepsilon_{t+j} + \beta_1 \cdot \varepsilon_{t+j-1} + \beta_2 \cdot \varepsilon_{t+j-2} + \cdots) = \sum_{i=0}^{\infty} \beta_{j+i} \cdot \varepsilon_{t-i} \tag{A.13}$$

Since changes in the trend are therefore unforecastable, this has the effect of decomposing the series into a random walk and a cyclical component, so that

$$y_t = \tau_t + c_t \tag{A.14}$$

where the trend is

$$\tau_t = \tau_{t-1} + e_t$$

and e_t is white noise.

To use the Beveridge-Nelson decomposition we must therefore: (1) Identify p and q in our ARIMA(p,1,q) model. (2) Identify the $\{\beta_j\}$ in equation 6. (3) Choose some large enough but finite value of s to approximate the limit in equation 8. (4) For all t and for $j = 1, \cdots, s$, calculate $E_t(e_{t+j})$ from equation 9. (5) Calculate the trend at time t as $y_t + E_t(\sum_{j=1}^{s} e_{t+j})$ and the cycle as y_t minus the trend.

Based on results for the full sample, we use an ARIMA(1,1,2), with parameters re-estimated by maximum likelihood methods before each recalculation of the trend.

A.6 The Blanchard-Quah Decomposition

The Blanchard-Quah measure of the output gap is based on a structural VAR identified via restrictions on the long-run effects of the structural shocks. Our implementation is identical to that of Cayen (2001), who uses a trivariate system including both output, CPI and yields on 3-month treasury bills. Lag lengths for the VAR are selected using corrected LR tests and a general-to-specific approach.

Table 1

Reliability of Alternative Output Gap Measures

Method	COR	AR	NS	NSR	OPSIGN
Linear Trend	0.87	0.93	0.50	1.36	0.53
Quadratic Trend	0.61	0.95	0.95	0.98	0.31
Breaking Trend	0.78	0.85	0.80	0.81	0.21
Hodrick-Prescott	0.50	0.92	1.10	1.10	0.38
Band Pass	0.69	0.78	0.73	0.81	0.32
Beveridge-Nelson	0.82	0.02	0.60	0.62	0.22
Blanchard-Quah	0.67	0.87	1.04	1.06	0.21
Watson	0.90	0.88	0.54	1.25	0.24
Harvey-Clark	0.88	0.88	0.61	0.64	0.13
Harvey-Jaeger	0.94	0.90	0.49	0.49	0.07
Kuttner	0.87	0.92	0.51	1.19	0.53
Gerlach-Smets	0.75	0.83	0.78	1.11	0.36

Notes: The table present summary measures of the reliability of real-time estimates of the output gap for 12 alternative methods of estimating the output gap. All statistics are for the 1969:1–1998:4 period. COR, denotes the correlation of the real-time and final estimates of the output gap. AR the first order serial correlation of the revision (the difference between the final and real-time series). NS indicates the ratio of the standard deviation of the revision and the standard deviation of the final estimate of the gap. NSR indicates the ratio of the root mean square of the revision and the standard deviation of the final estimate of the gap. OPSIGN indicates the frequency with which the real-time and final gap estimates have opposite signs.

Table 2

RMSE of Forecasts—In Sample

Method	1969-1998	1969-1983	1984-1998
Linear Trend	1.601	1.953	1.137
Quadratic Trend	1.629	1.971	1.183
Breaking Trend	1.741	2.127	1.230
Hodrick-Prescott	1.662	1.885	1.399
Band Pass	1.765	2.135	1.284
Beveridge-Nelson	1.746	2.085	1.315
Blanchard-Quah	1.742	2.067	1.334
Watson	1.623	1.972	1.165
Harvey-Clark	1.728	2.077	1.279
Harvey-Jaeger	1.798	2.048	1.503
Kuttner	1.550	1.902	1.079
Gerlach-Smets	1.470	1.747	1.119
AR	1.912	2.340	1.344
Real Growth	1.750	2.078	1.337
Nominal Growth	1.550	1.737	1.332

Notes: The entries show the RMSE of the inflation forecast from equation (1). The first twelve rows show results using alternative output gaps. The AR forecast is univariate, and the last two rows show the forecasts based on real and nominal growth instead of the gaps.

Table 3

RMSE Relative to AR—In Sample

Method	1969-1998	1969-1983	1984-1998
Linear Trend	0.838**	0.835**	0.846**
Quadratic Trend	0.852**	0.843**	0.880
Breaking Trend	0.910**	0.909**	0.915**
Hodrick-Prescott	0.869*	0.805**	1.041
Band Pass	0.923**	0.913**	0.956
Beveridge-Nelson	0.913**	0.891**	0.978
Blanchard-Quah	0.911**	0.883**	0.992
Watson	0.849**	0.843**	0.867*
Harvey-Clark	0.904**	0.888**	0.952
Harvey-Jaeger	0.941	0.875**	1.119
Kuttner	0.810**	0.813**	0.803**
Gerlach-Smets	0.769**	0.747**	0.833*
Real Growth	0.915**	0.888**	0.995
Nominal Growth	0.811**	0.743**	0.991

Notes: The entries show the ratio of the RMSE of the inflation forecast based on the method shown and the RMSE of the AR forecast. * and ** indicate that the improvement in forecasts relative to the AR forecast, as measured by the modified Diebold-Marianno test statistic, are statistically significant at the 10 and 5 percent levels, respectively.

Table 4

RMSE of Forecasts—Real Time

Method	1969-1998	1969-1983	1984-1998
Linear Trend	2.341	2.821	1.719
Quadratic Trend	2.390	2.878	1.761
Breaking Trend	2.377	2.936	1.622
Hodrick-Prescott	2.333	2.622	1.998
Band Pass	2.244	2.634	1.763
Beveridge-Nelson	2.185	2.681	1.522
Blanchard-Quah	2.483	3.007	1.798
Watson	2.214	2.764	1.453
Harvey-Clark	2.592	3.263	1.648
Harvey-Jaeger	2.111	2.316	1.880
Kuttner	2.254	2.807	1.495
Gerlach-Smets	2.242	2.810	1.451
AR	2.308	2.869	1.540
Real Growth	2.153	2.587	1.596
Nominal Growth	1.946	2.252	1.576

Notes: See notes to Tables 2 and 3.

Table 5

RMSE Relative to AR—Real Time

Method	1969-1998	1969-1983	1984-1998
Linear Trend	1.014	0.983	1.116
Quadratic Trend	1.036	1.003	1.144
Breaking Trend	1.030	1.023	1.053
Hodrick-Prescott	1.011	0.914	1.297
Band Pass	0.973	0.918	1.144
Beveridge-Nelson	0.947	0.935	0.988
Blanchard-Quah	1.076	1.048	1.168
Watson	0.959	0.964	0.944
Harvey-Clark	1.123	1.138	1.070
Harvey-Jaeger	0.915	0.807	1.221
Kuttner	0.977	0.979	0.971
Gerlach-Smets	0.971	0.980	0.942
Real Growth	0.933	0.902*	1.036
Nominal Growth	0.843**	0.785**	1.023

Notes: See notes to Tables 2 and 3.

Table 6

Relative RMSE: 1969-1998—Real Time

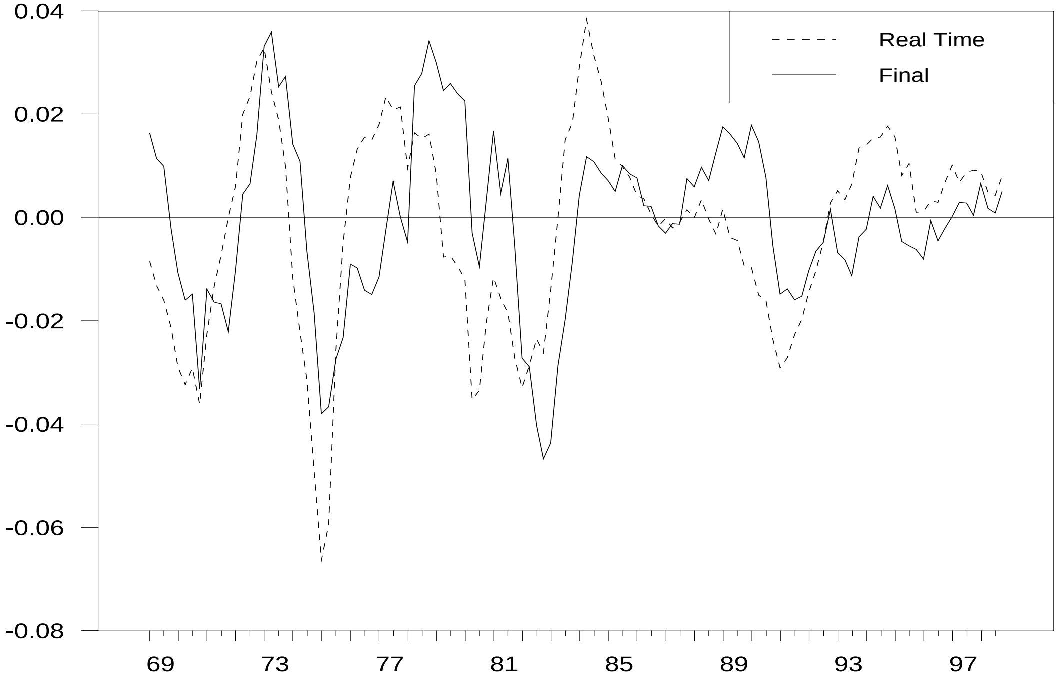
Method	AR	Real Growth	Nominal Growth
Linear Trend	1.014	1.087	1.203
Quadratic Trend	1.036	1.110	1.228
Breaking Trend	1.030	1.104	1.221
Hodrick-Prescott	1.011	1.084	1.199
Band Pass	0.973	1.042	1.153
Beveridge-Nelson	0.947	1.014	1.122
Blanchard-Quah	1.076	1.153	1.276
Watson	0.959	1.028	1.137
Harvey-Clark	1.123	1.204	1.332
Harvey-Jaeger	0.915	0.980	1.085
Kuttner	0.977	1.047	1.158
Gerlach-Smets	0.971	1.041	1.152
AR		1.072	1.186
Real Growth	0.933		1.106
Nominal Growth	0.843	0.904	

Notes: Each entry denotes the ratio of the RMSE of the inflation forecast based on the methods shown in the corresponding row to the RMES based on the method shown in the corresponding column.

Figure 1

Real-Time and Final Hodrick-Prescott Output Gap

One Quarter



Four-Quarter Moving Average

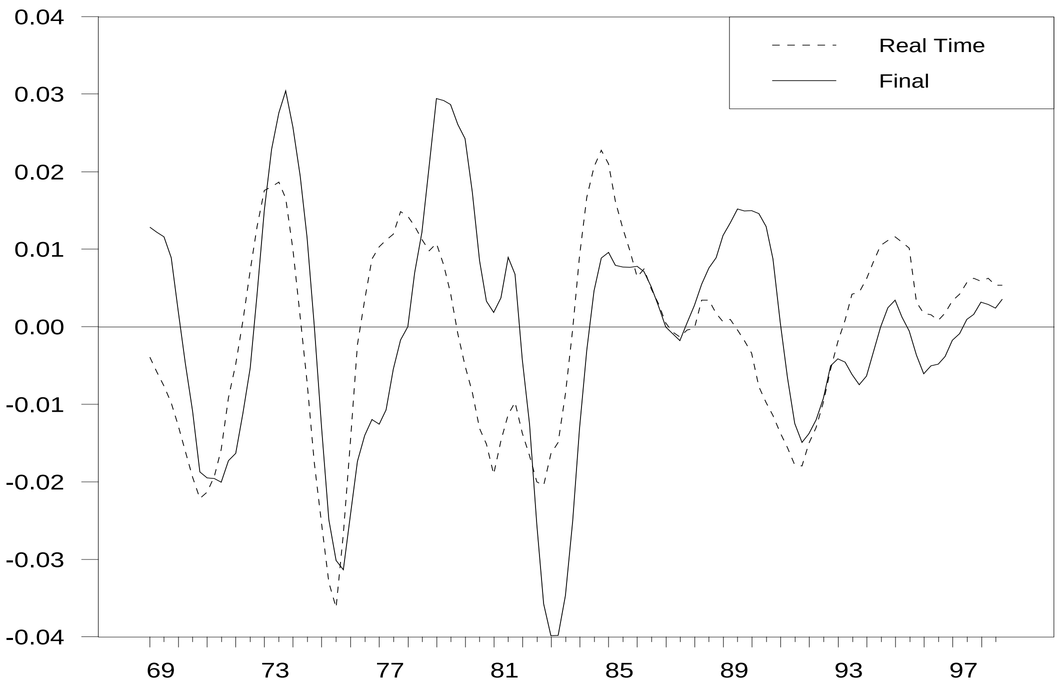
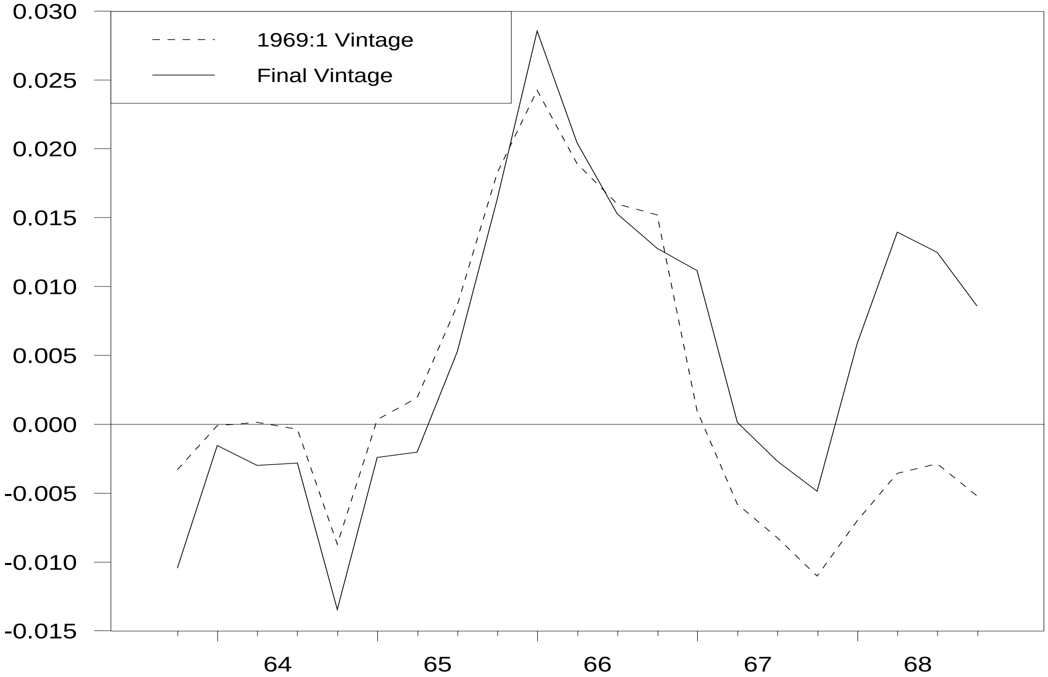


Figure 2

Historical Vintages and Final Hodrick-Prescott Output Gap

1969Q1 Vintage



1996Q1 Vintage

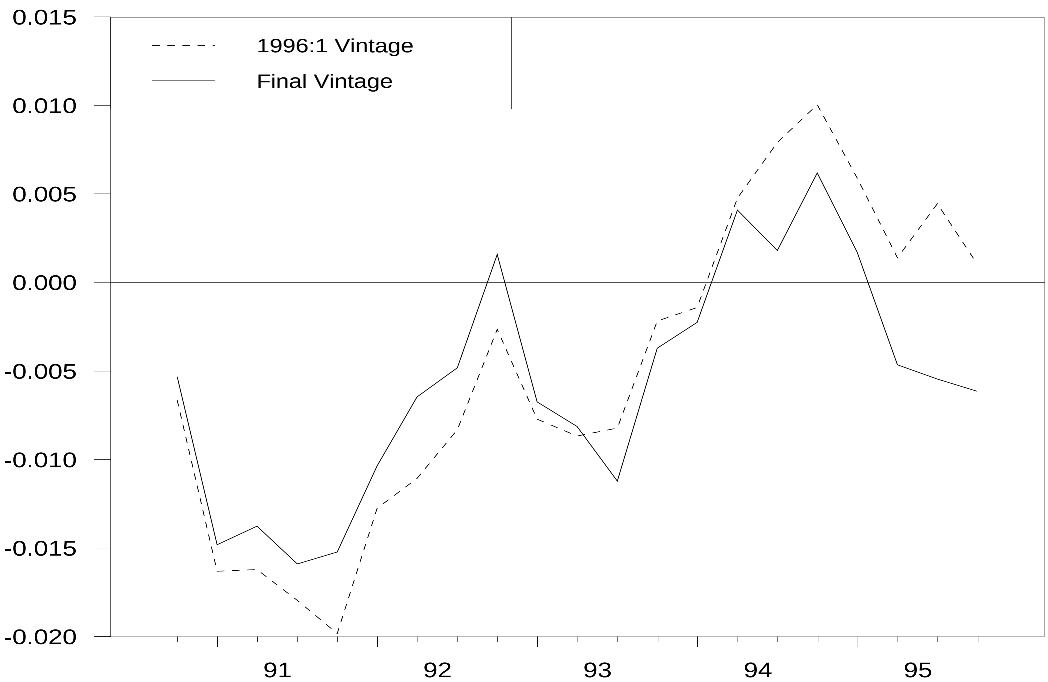
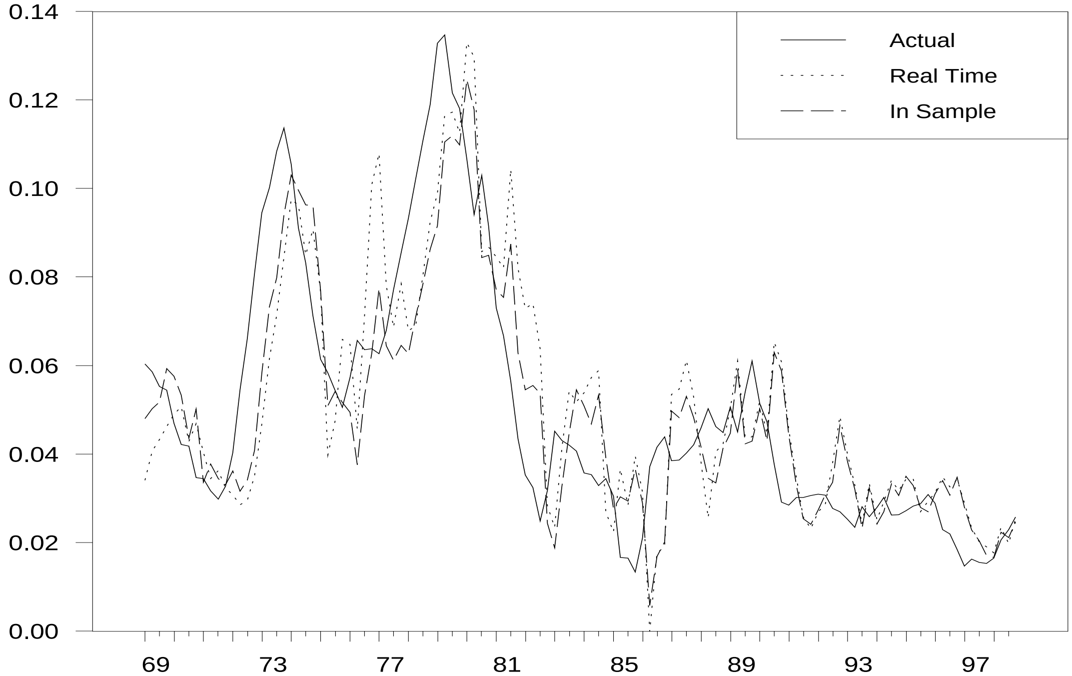


Figure 3

Univariate Inflation Forecasts

Forecasts and Realized Inflation



Forecast Errors

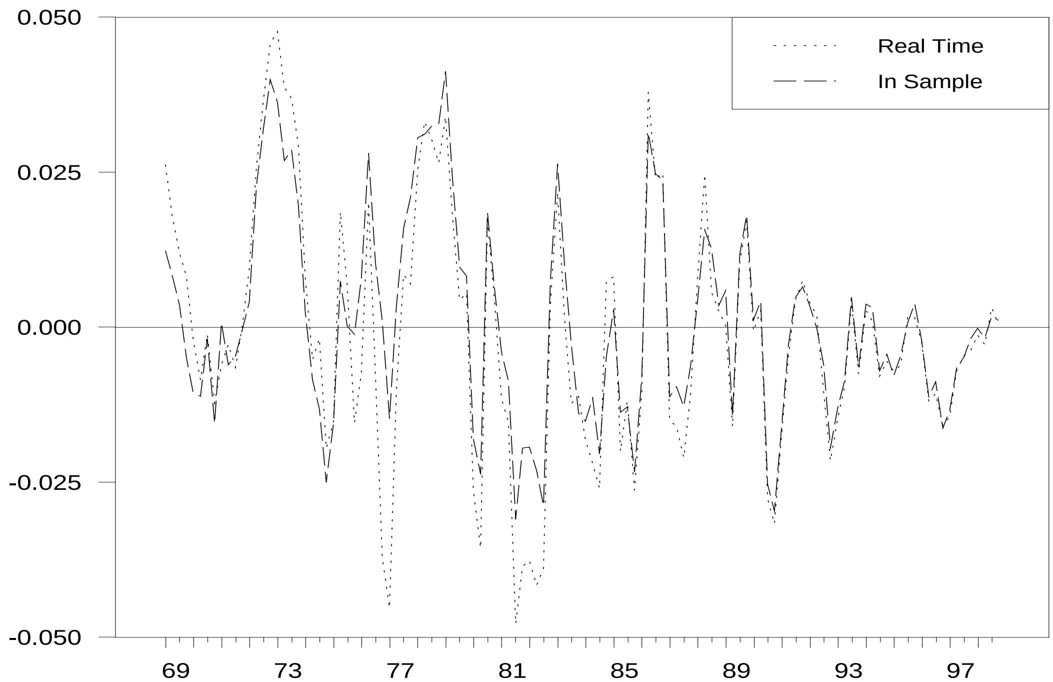
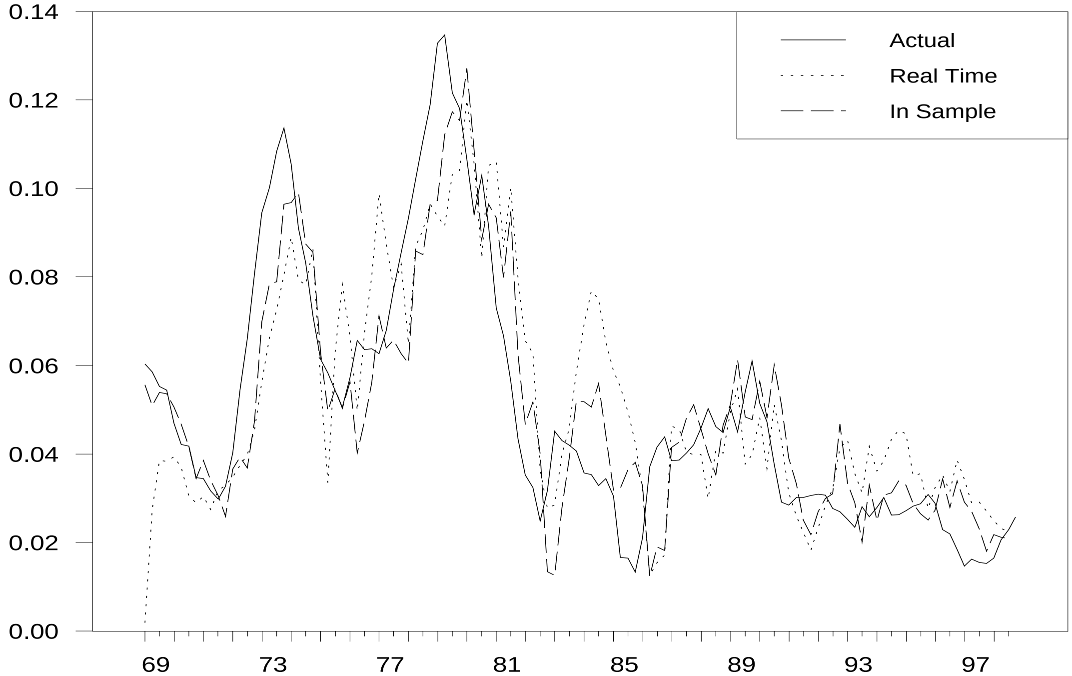


Figure 4

Inflation Forecasts and Errors with the Hodrick-Prescott Output Gap

Forecasts and Realized Inflation



Forecasts Errors