What Can We Learn From Comprehensive Data Revisions for Forecasting Inflation? Some U.S. Evidence*

Pierre L. Siklos
Wilfrid Laurier University
and
Viessman Research Centre
Waterloo, ON
Canada

July 2007

Abstract

The empirical properties of benchmark revisions to key U.S. macroeconomic aggregates are examined. News versus noise impact of revisions is interpreted via the cointegration property of successive benchmark revisions. Cointegration breaks down in the last two years before a benchmark revision. Hence, we conclude that there is some information content in benchmark revisions. This last point is illustrated by reporting that inflation forecasts could be improved by the addition of a time series that reflects benchmark revisions to real GDP. Standard backward- and forward-looking Phillips curves are used to explore the statistical significance of benchmark revisions.

JEL classifications: E01, E31, C53

Keywords: Cointegration breakdown; Real-time data; Phillips curve

* Siklos (corresponding author): Department of Economics, Wilfrid Laurier University, 75 University Ave., Waterloo, ON, Canada, N2L 3C5 (psiklos@wlu.ca).
1. Introduction

There is a resurgence of interest in the role played by data revisions in policy analysis. For example, it has become apparent that the appearance of relatively poor conduct in monetary policy during the 1970s and 1980s is, in hindsight at least, partly attributable to differences between data available to policy members at the time decisions were taken and subsequently revised data. Orphanides (2001) is an oft-cited study revealing how reliance on final revised data in econometric studies can lead to a form of historical revisionism (also see, e.g., Runkle, 1998; Orphanides and van Norden, 2005). Since then there has been a veritable outpouring of academic research in this area. The collection of articles edited by Herrmann, Orphanides, and Siklos (2005) is a recent addition to the literature that considers the international experience with real-time data. While the potential consequences of relying on different vintages of data has been known for some time, it is only recently that new data sets have been made widely available for research, leading to a revival of interest in capturing and storing data (and models) that policymakers use in real time.¹

An important feature of the revision process takes place on a regular basis in many countries, including the U.S. Benchmark, also known as comprehensive, revisions to the National Income and Product Accounts (NIPA) that update the data in light of the evolution of the U.S. economy, reflect changes in definitions for some series, include newly available or revised data, and are introduced roughly every five years.² If revisions to the data are informative, one would want to know whether such releases are potentially useful to forecasters. Indeed, one can well imagine that new vintages of data produced quinquennially could be used as a means of verifying whether past forecasts could have been improved by

¹ Dean Croushore has done a great service to the profession not only by maintaining a web page that updates past and present research dealing with real-time data but also by pioneering the development of real-time data sets for the U.S. economy. See oncampus.richmond.edu/~dcrousho/docs/realtime_lit.pdf, which periodically updates the ever-expanding literature on real time data, www.philadelphiafed.org/econ/forecast/reaindex.html for updates to U.S. real time series, and Croushore and Stark (2005). The movement to develop real-time data sets has become an international phenomenon; e.g., http://www.eabcn.org/ which provides a real-time data set for the euro area. Finally, the Federal Reserve Bank of St. Louis has also begun to archive vintages of data covering a wide variety of time series at research.stlouisfed.org/fred2/vintageseries/.

² A good source for details about these, and other revisions to the data, can be found at the Bureau of Economic Analysis’s website of the U.S. Department of Commerce (www.bea.gov/beahome.html). Also see Ritter (2000). It appears that National Accounts in the U.S. are revised more frequently than elsewhere in the world, where revisions tend to be undertaken about once a decade.
incorporating information about the impact of revisions on successive vintages of data. Moreover, depending on the time-series properties of successive revisions, these may also have implications for the manner in which certain popular macro models used in policy analysis should be estimated. Croushore and Stark (2002) were surprised to find that the forecasting performance of a model that relies on the current vintage of data is no different than if the researcher had generated a forecast using an earlier vintage. Bernanke and Boivin (2003) reach essentially the same conclusion in a different setting. However, the results of such studies may be sensitive to the choice of models, and the manner in which revisions are thought to evolve over time.

The aim of this paper is to revisit the time-series properties of such revisions in order to address some of the issues raised above. Siklos (1996) relies on the cointegration methodology to explore the properties of data revisions. Cointegration is the property which captures the notion that certain economic time series are attracted to each other. Other than Golinelli and Parigi (2005) and Siklos (1996), the literature has tended to eschew the cointegration approach to the analysis of data revisions. Instead, several authors (e.g., Faust, Rogers, and Wright, 2005; Garratt and Vahey, 2006) compare growth rates of time series that are frequently revised across vintages, as opposed to examining the properties of revisions in the levels. It is not immediately clear that one approach offers an advantage over another. Comparing revisions in growth rates can mitigate some of the minor changes in statistical or reporting procedures that could contaminate the levels data. Aruoba (2005) concludes that revisions are not well-behaved in that they are biased and errors can be large. Cointegration, however, offers the opportunity to test whether certain vintages are attracted to each other. If we do not expect vintages to be cointegrated, perhaps because benchmark revisions reflect changes in economic structure over time that upset what normally would be stationary differences between them, it might be useful to test whether this possibility describes the revision process. Currently, the investigator simply assumes that the culprit must be the nature of the

---

3 Implicitly, we consider only the case where the levels of each series are integrated of order one, or \( I(1) \), so that first differencing is usually sufficient to render such series \( I(0) \).
revision process.\textsuperscript{4} However, what if the cointegration property is restored in the presence of a break? Moreover, what if the break is detected nearer the time a benchmark revision is being introduced? It is possible that this outcome reflects deterioration in the ability of key time series to track the evolution of the economy over time. In any event, the cointegration breakdown could be informative about changes in the economy that might serve as a useful input in a forecasting exercise.

In the following section we adopt the cointegration approach to investigate the statistical properties of benchmark revisions. In particular, it is found that the presence or failure of the cointegration property can take place toward the end of a sample. End-of-sample cointegration breakdown tests, recently introduced by Andrews (2003) and Andrews and Kim (2006), are used to make the point. In other words, the cointegration property between vintages is restored once we allow for a break in the “long-run” relationship that exists between vintages of data that are not temporally close to each other.

Next, we explore some potential sources of any breakdown in cointegration. We find that benchmark revisions can be correlated with changes in asset prices, such as equity returns, housing price inflation, and interest rate changes. What remains unclear is whether asset prices are influenced by the information content in the quinquennial revisions or vice-versa. The paper does not take a stand on the causal relationship between benchmark revisions and asset price behavior. It merely suggests the possibility that the time-series properties of benchmark revisions might capture some of the influences that are reflected in the behavior of asset prices.\textsuperscript{5} For example, if the results of this paper reflect the fact that the U.S. economy evolves rapidly over time, a better understanding of the process of technological change, and empirical models that are better able of capturing such effects, should be further encouraged. The present paper, however, does not explicitly address this question. It merely points out the potential usefulness of comprehensive data revisions for forecasting purposes. In particular, the findings suggest that

\textsuperscript{4} This is essentially the argument made by Gollinelli and Parigi (2005). However, they do not consider the possibility that the presence (or absence) of cointegration could be influenced by a structural break. As we shall see, there are at least two sources of change that lead to a rejection of the cointegration hypothesis, namely changes in definitions and the rebasing of some series.

\textsuperscript{5} Another potentially important source of the breakdown in cointegration could be changes in trend productivity growth. This possibility is not investigated here; however, see van Norden (2005).
quinquennial revisions may either be treated as an error-correction type series, or possibly as an instrument in some forward-looking model, or both. We illustrate the point by estimating simple Phillips curves models of the kind frequently reported in the literature. We also conclude that some published forecasts, either by private or public agencies, could have been improved using information contained in benchmark revisions.

The rest of the paper is organized as follows. The next section outlines methodological issues. Section 3 briefly describes the data, and presents some stylized facts, as well as cointegration and cointegration breakdown test results. Section 4 explores empirically the implications of our findings for the Phillips curve. Section 5 concludes.

2. Methodological Issues

A fairly standard way of analyzing the time-series properties of revisions to data is to determine what fraction represents “news,” that is, potentially new, and likely useful, information, versus “noise.” The latter represents that portion of the chosen vintage that is uncorrelated with true values (e.g., final revised data). If all available information is incorporated in one particular vintage (e.g., preliminary releases of GDP data) by the statistical agency responsible for publishing the data, then “news” is information that arrives after the data are released. As a result, earlier vintages are optimal forecasts of subsequent revisions of the data. More formally, we can express the relationship between a vintage \((v)\) and a subsequent one, which we refer to, for the time being, as the “final” vintage \((f)\), as follows:

\[
X^v_t = X^f_t + \varepsilon_t, \tag{1}
\]

where \(X^v_t\) is a time series for, say, GDP, released temporally prior to the final release of the same series, denoted \(X^f_t\). In this setup, and denoting \(\rho\) as the correlation coefficient, if \(\rho(X^f_t, \varepsilon_t) = 0\), then revisions are pure noise, since final revised data are uncorrelated with the error term, while in the event that

---

6 This presumes, of course, that the goal of the statistical agency is to minimize errors or biases, an assumption that is both sensible and, under the present circumstances, reasonable.
\( \rho(X_t^v, \varepsilon_t) = 0 \), this would indicate support for the news view, since the earlier vintage is now uncorrelated with the error term. Thus, the particular vintage is completely uninformative about the final estimate. Put differently, revisions contain news if the revisions are uncorrelated with an earlier vintage. In contrast, the noise hypothesis implies that revisions are uncorrelated with final data. Intermediate cases, where elements of news and noise coexist, are also possible and empirically more likely. The foregoing dichotomy is the one introduced by Mankiw, Runkle, and Shapiro (1984), Mankiw and Shapiro (1986), and used by many others since with some modifications (e.g., Faust, Rogers, and Wright, 2005).

Since (1) is not specific about how temporally close or distant \( X_t^v \) is from \( X_t^f \), there is also the potential for \( \rho \) to be a function of \( v - f \). If we define revisions employing the following expression,

\[
R_t = X_t^v - X_t^f ,
\]

then a regression of the form,

\[
R_t = \alpha + \beta X_t^v + u_t ,
\]

also known as a Mincer-Zarnowitz regression, is used to test efficiency which requires non-rejection of the null hypothesis that \( H_0 : \alpha = \beta = 0 \). In other words, if \( X_t^v, X_t^f \sim I(1) \), then under the null \( R_t \sim I(0) \). Since the focus of this paper is on the information content of benchmark revisions, a few words concerning the underlying statistical model of benchmark revisions is in order. Croushore and Evans (2006) specify the following statistical model for benchmark revisions:

\[
X_t^{v+s} = X_t^v + \eta_t^{v+s},
\]

where \( v+s \) indicates that we are only concerned with revisions \( s \) periods apart, \( \eta_t^{v+s} \) is a pure measurement error, and \( s \geq T \). Therefore, beyond some threshold \( T \) (chosen, not “estimated”), revisions are assumed to be completely random.\(^7\) Now, while specification (4) may be correct, there is little compelling reason to justify it, except perhaps that this model is consistent with the “surprising” result

---

\(^7\) Within benchmark revisions, however, measurement errors may be serially correlated. Croushore and Stark (2002) focus on the vintage that immediately precedes a benchmark revision, and not the first vintage of a new benchmark revision.
reported in Croushore and Stark (2002). Alternatively, one could argue that successive benchmarks correct past errors as well as providing improvements in the quality of previous estimates. Nevertheless, as Ritter (2000) points out, although the quality of NIPA data is high, those most acquainted with the data collection process suggest that the reliability of, for example, GDP data is an ongoing issue and presumably not restricted to data published within benchmark revisions. Therefore, the paper posits that the true value of $X_t$, namely $X_t^r$, is possibly obscured by important changes in the structure of the economy that are apparent nearer the time of benchmark revisions. Why a structural break should occur as the date of a benchmark revision approaches is unclear. Perhaps experience has taught those who are responsible for producing these data that there are features in an evolving economy that are increasingly poorly captured by the existing definitions and methods, and that five years is approximately adequate to deal with the necessary revisions. Of course, since benchmark revisions entail considerable costs, there must exist implicitly a trade-off between an appropriate time delay between benchmark revisions and the need to avoid deterioration in the data quality beyond some threshold. The tests and empirical results that follow do not explicitly incorporate the foregoing considerations, as this would require a formal model of the data revision process, and this is beyond the scope of this paper.

The approach adopted in this paper is related to the earlier literature, but the testing strategy employed here is different. Since we are interested in revisions that, we presume, reflect structural or, possibly, longer run influences on the data, the relationship between $X_t^r$ and $X_t^f$ potentially exists via the cointegrating property. Under this view, $R_t$ is a stationary process or $I(0)$, that is, it does not contain a unit root, if the benchmark revisions contain little or no statistically meaningful information, while significant structural shifts in the economy would lead to the non-rejection of the null of a unit root. More formally, we can write

$$X_t^r = \theta_0 + \theta_1 X_t^f + u_t.$$  

(5)

---

8 While the present paper focuses on NIPA revisions, the time-series properties of the data are also likely to be affected by periodic censuses.
As in the Mincer-Zarnowitz formulation, non-rejection of the null hypothesis that \( \theta_0 = 0, \theta_1 = 1 \) reduces (5) to (1). The residual-based test of the cointegration property requires only that \( u_t \sim I(0) \), that is, the residuals should be integrated of order zero, while the additional restrictions that \( \theta_0 = 0, \theta_1 = 1 \) describe the cointegrating vector between \( X_t^v \) and \( X_t^f \). Of course, both \( X_t^v \) and \( X_t^f \) must be \( I(1) \). Equation (5) suggests that a bivariate relationship between any \( X_t^v \) and \( X_t^f \) is the only one of interest. It is conceivable, however, that successive benchmark revisions are related to each other, if only because subsequent vintages presumably reflect learning and other improvements in the data gathering apparatus the statistical agency has accumulated over time. Hence, if we denote the most recent vintage as \( X_t^f \) and earlier benchmark revisions as \( X_t^v \), where \( i \) denotes the particular year a benchmark revision is published (every five years in the present case), then it is conceivable that the appropriate cointegration test is between the set of all available vintages around benchmark revisions. We can also write this cointegrating relationship in a VAR format (see Johansen, 1995):

\[
\begin{pmatrix}
X_t^f \\
X_t^v \\
\vdots \\
X_t^v \end{pmatrix} = \Phi(L)x_t + \begin{pmatrix}
e_t^f \\
e_t^v \\
\vdots \\
e_t^v \end{pmatrix},
\]

(6)

where \( X_t^v, \ldots, X_t^v \) are the different benchmark vintage that temporally precede \( X_t^f \), and \( x_t \) is a vector containing lags of \( X_t^v, \ldots, X_t^v, X_t^f \).

If there is cointegration, this will be reflected in the rank of \( \pi = \alpha \beta' \) in the error-correction form,

\[
\Delta x_t^v = \pi x_{t-1}^v + \sum_{i=1}^{\mu-1} \pi_i \Delta x_{t-i}^v + \epsilon_t,
\]

(7)

where \( \pi \) is the product of the transpose of the matrix of cointegrating vectors \( \beta' \) and the matrix of speed-of-adjustment parameters \( \alpha \). Clearly, the breakdown in the relationship between vintages can originate from more than one source, further complicating the analysis of benchmark revisions in the multivariate setting. While (6) and (7) represent the most general form of the potential cointegrating
relationship between successive vintages, in what follows we focus on the pair of vintages $X_t^v$ and $X_t^f$. This approach allows us to explore in a simpler framework the source of a potential breakdown in a cointegrating relationship between vintages of data. In addition, it seems likely that, over several benchmark revisions, the information content in the data revisions would be contaminated by possibly more than one structural break. The procedure used here permits the identification of a single structural break.

It is well known that cointegration can breakdown for a number of reasons, including the presence of a structural break (Siklos and Granger, 1997). Indeed, depending upon the exact timing of the benchmark revisions, the breakdown in cointegration can conceivably take place anytime. However, if final revised data contain the “truth” about underlying time series then revision errors are likely to be greatest for the earliest vintages. In other words, it is likely that, other things equal, the earlier the vintage, the smaller is the predictive content for final revised data. Moreover, it is also conceivable that loss of predictability may be largest for more recent observations than for more temporally distant observations. Hence, the breakdown of the cointegration property, if it occurs at all, is arguably more likely to occur toward the end of a particular sample. Existing tests of breaks in a cointegrating relationship (e.g., Gregory and Hansen, 1996; Hansen and Johansen, 1999) are not well suited to handle such a situation. Andrews (1993) considers the statistical properties of conventional tests, such as the Chow and Wald tests, when one searches for a break overall possible dates in a sample. The proposed test statistic, however, includes a “trim factor,” as the asymptotic theory requires it. Generally, the convention is to set the trim factor at 15%, meaning that conventional tests cannot be used to conduct inferences about structural breaks for the first and last 15% of each sample. By contrast, the recently proposed tests of Andrews (2003) and Andrews and Kim (2006) are the appropriate ones under the present circumstances, since they can be applied to the end of a particular sample.9

---

9 As we shall see, while the Andrews and Kim (2006) test solves one problem it leaves a question unanswered, as it is a residual-based test. When a test based on (6) is conducted there may be more than one cointegrating relationship and the issue then is whether there are potentially several breakdowns in the cointegrating relationship. In this paper we are only interested in whether there is at least one failure to find cointegration. Presumably, if there was a failure
Assume that a sample of data is split at $t = T$, where $T + m$ is the total number of observations in the sample. If we write the cointegrating relationship as in (5) and there is a breakdown in cointegration in the sub-sample $[T+1, T+m]$, then the linear relationship between $X_t^r$ and $X_t^f$ can be written as

$$y_t = \begin{cases} x'_t \theta_0 + u_t, & t = 1, \ldots, T \\ x'_t \theta + u_t, & t = T + 1, \ldots, T + m \end{cases}$$

(8)

The regressor $x'_t = X_t^f$ is the case depicted in (5) with $y_t = X_t^r$, but other variables can be included, whether they are $I(1)$, stationary, or incorporate a deterministic component. As explained in Andrews and Kim (2006), the breakdown of cointegration can occur either because $\theta$ changes or because $u_t$ ceases to be $I(0)$, as would be true if the cointegration property holds, and becomes $I(1)$ if cointegration ceases to exist. If the vintages are cointegrated and their relationship is stable, this is akin to testing the null hypothesis that $H_0: \theta_t = \theta_0$ for $t = T + 1, \ldots, T + m$ with $u_t \sim I(0)$ or that $u_t \sim I(0)$ for $t = T + 1, \ldots, T + m$. The alternative hypothesis is that $\theta_t \neq \theta_0$, for some $t = T + 1, \ldots, T + m$ or that the distribution of $\{u_{T+1}, \ldots, u_{T+m}\}$ becomes non-stationary. Andrews and Kim (2006) develop test statistics of the Chow variety, and the choice of test statistics is sensitive to the location of the split in the sample; we focus on their $P_C$ and $R_C$ statistics. Andrews and Kim (2006) conduct an extensive series Monte Carlo simulations analyzing the power and size of the test statistics and conclude that the tests based on the $P_C$ and $R_C$ statistics are recommended.\(^{10}\) Critical values are determined via parametric sub-sampling.\(^{11}\)

---

\(^{9}\) If $\theta$ in (8) is estimated over the sample $t = 1, \ldots, T$, then we denote the estimate by $\hat{\theta}_{1,T}$. Andrews and Kim (2006) suggest that the estimators $\hat{\theta}_{1-(T-m/2)}$ and $\hat{\theta}_{1-(T+m)}$ have relatively better statistical properties. Moreover, tests relying on $\hat{\theta}_{1-(T+m/2)}$ were found to have the most favorable power properties against either of the alternatives of a shift in the parameters in the cointegrating regression or a change in the distribution of $u_t$ from $I(0)$ to $I(1)$. Test results given in the next section are based on critical values for this estimator.

\(^{11}\) Essentially, this means that the distribution of the test statistic is derived from sequentially applying the test to a stable sample; see Andrews (2003).
To summarize, we would normally expect $u_t$ in (5) to at least be stationary, that is, $u_t \sim I(0)$.$^{12}$

Recall that, for the purposes of the exercise conducted here, $u_t$ represents the residuals between different pairs of vintages with the basis of comparison always relative to the final revised data. The claim made in the paper is that the breakdown of cointegration, if it occurs, is a feature of the end of the particular sample in question. Moreover, as will be demonstrated empirically in the following section, the finding of a breakdown in cointegration can take place around the time of what appears in hindsight to be important benchmark revisions that portend significant structural changes in the U.S. economy. As a result, certain benchmark revisions may contain some useful predictive content for key macroeconomic time series.

Since no test is definitive, we augment our analysis of benchmark revisions by asking whether these could have usefully been employed to improve inflation forecasts, and whether any information content in such revisions can be traced to variables that might signal important structural changes in the economy, such as, for example, those incorporated in the behavior of asset prices.

3. Data and Empirical Evidence

3.1. Data

Quarterly data for real GNP/GDP and real total personal consumption expenditure were obtained from the Federal Reserve Bank of Philadelphia’s real-time data website (www.philadelphiafed.org/econ/forecast/reaindex.html; also see Croushore and Stark, 2002). In the results reported below, all data are seasonally adjusted at the source and it is possible that this too imparts an additional distortion that is not directly controlled for (see, e.g., Ghysels, Granger, and Siklos, 1996). The present paper focuses on eight benchmark revisions in 1966, 1971, 1976, 1981, 1986, 1992, 1996, and 2001, all occurring in February of the years listed. These revisions are assumed to have occurred in the first quarter of the benchmark revision year. The vintage 2004Q1 (or, more precisely, February 2004)

$^{12}$ Since the metric is, in part, whether $u_t \sim I(0)$ or not, this condition is weaker than the requirement of white noise. Hence, as argued below, findings based on the cointegration and cointegration breakdown exercises are not sufficient. Other tests should also be considered prior to reaching more definite conclusions about the information content of successive benchmark revisions.
represents the “final” vintage in our data set (i.e., this vintage proxies $X_f'$). Since the arguments in this paper rely on the notion that successive benchmark revisions can be informative, readers should be made aware of the fact that several of these benchmark revisions reflect not only changes in definitions, and the switch from GNP to GDP in 1992, but also changes in the base year (in January 1976, in December 1985, in November 1991, and in January 1996). For the forecasting exercise reported below, only the 1996 base year change is an issue.\footnote{As in Croushore and Stark (2002), base year changes can be handled by subtracting the mean difference over the common sample for pairs of vintages being compared. In addition, until 1991, vintages rely on the fixed-weight index. Subsequently, the chain-weighted index is employed. It is also unclear, for the purposes of this paper, whether base year changes should be treated differently from other definitional changes. There are a variety of reasons for such changes, and not all reflect re-basing. It should also be noted that re-basing is also done from time to time to account for changing patterns in consumption spending by households.}

Although the tests described in the previous section were performed on all series considered, the discussion below focuses on real output and a proxy for the output gap. In the case of the output gap we chose to apply an Hodrick-Prescott (HP) filter (with smoothing parameter 1600) to the log-level of real GDP. Standard HP detrending has well-known drawbacks, such as the end-point problem, and the possibility of inducing spurious cycles, but most researchers employ this filter to create a proxy for the output gap.\footnote{Some experiments that rely on the measure of potential output data from the Congressional Budget Office (www.cbo.gov), a widely used time series, did not change any of our conclusions.} In the case of inflation, we used the data from the latest available vintage, as it is a fairly well accepted proposition that revisions to this series are quite small (see, e.g., Kozicki, 2004).\footnote{Again, some experimentation with different vintages of the CPI confirms this to be the case.} Therefore, we follow others in the literature and concentrate on the properties of revisions to output and consumption data.

Turning to the analysis of the impact of informative benchmark revisions on forecasts of inflation, we obtained data from a variety of sources. They include monthly forecasts from The Economist and Consensus Economics. Data from The Economist were collected from the hard copy version of this publication, as were Consensus forecasts (www.consensus.com). Monthly forecasts were averaged to obtain data at the quarterly frequency. We also used semi-annual forecasts from the OECD, and these data
were also converted to the quarterly frequency using cubic-match last interpolation.\textsuperscript{16} Survey of Professional Forecasters (SPF) and Livingstone Survey (LS) forecasts are available from the Federal Reserve Bank of Philadelphia, while OECD forecasts were collected from successive sources of its Economic Outlook publication (www.oecd.org). Available samples for these series vary. For example, OECD forecasts are available since 1960, Consensus and Economist forecasts only begin in 1990, while whereas SPF data begin in 1981.\textsuperscript{17} Clearly then, there are limitations to the number of benchmark revisions that can be examined against some of the published forecasts. Consequently, we augment our analysis of the potential impact of benchmark revisions by estimating a variety of backward- and forward-looking Phillips curves. The manner in which these Phillips curves are specified is outlined in greater detail below. The objective is to determine whether inflation forecasts could have been improved by the addition of benchmark revisions to real GDP in the estimated specifications.\textsuperscript{18}

\subsection*{3.2. Some Stylized Facts}

Figure 1 plots revisions to the log of real GDP for the eight benchmark revisions considered here. The plots reveal three striking features in the data (also see Croushore and Stark, 2002). First, the series appear to be non-stationary, so that we can expect the null hypothesis of no cointegration to be valid. Table 1 generally confirms this result for $R_t$ at least based on an augmented Dickey-Fuller test. Revisions to the output gap and real GDP growth are usually $I(0)$, but the evidence for real consumption growth is mixed. Interestingly, the finding that revisions in the log change of the series for real personal consumption expenditure and real GDP are $I(0)$, while the same is not generally true of revisions in the levels, helps explain why Mankiw and Shapiro (1986), as well as several other authors since, resort to

\textsuperscript{16} This technique refers to fitting a cubic function to fill the gap between one observation and the next one at one sampling frequency in order to create hypothetical observations at a higher frequency.

\textsuperscript{17} To conserve space, additional details about these forecasts are not discussed. In addition to the sources already listed, readers are referred to Siklos (2002) and Siklos and Wohar (2006), where these forecasts are extensively employed.

\textsuperscript{18} Also considered was the impact of revisions in real personal consumption expenditures and money growth. Results using the former series are comparable to ones reported here using real GDP, while money growth revisions did not improve inflation forecasts at all. To conserve space, the relevant results are not shown.
analyzing revisions in growth rates as opposed to the levels. Second, there are often large revisions near the end of each sample for most of the pairs shown in the figure. As a result, the apparent non-stationarity in benchmark revisions may be confined to the last few years of data. To highlight this stylized feature of the data, Figure 1 includes vertical bars which identifies the location of the last observation for each vintage shown. The non-stationarity in $R_t$ reported above may in fact be more of a feature of the end of each sample. The large end-of-sample revisions are more noticeable for some vintages than for others (e.g., 1966, 1976, and 2001). It is arguable whether some of the vintages are historically more important than others. For example, it was around 2001 that productivity improvements over the previous decade were believed to have been confirmed by the data.\(^{19}\) This will become even more apparent, as we shall see when we examine first differences in the log of real GDP.

![FIGURE 1 ABOUT HERE](image-url)

![TABLE 1 ABOUT HERE](image-url)

Figure 2 plots several estimates of $R_t \delta$ for the output gap. Here, $X_t^{\delta}$ is an estimate of the output gap for the 2004Q1 vintage, while $X_t^{\delta_v}$ are the vintages associated with benchmark revisions that took place in the years shown in the legend. By construction, the output gap is expected to be $\delta(0)$ for the full sample over which the HP filter is applied.\(^{20}\) However, due to the end-point problem with HP filters, this

\(^{19}\) Meyer (2004) describes in vivid detail both the skepticism shared by many about Alan Greenspan’s view that productivity gains in the U.S. economy were real and the subsequent confirmation of this development. A possible difficulty with this interpretation is that 2001 also marks the year the U.S. went through a brief recession (http://www.nber.org/cycles.html).

\(^{20}\) I did not experiment extensively with one-sided HP filters nor with padding the data at the end of the sample. However, the one-sided filter did not appear to have a noticeable impact on the overall time-series properties of the data.
does not guarantee that the \( I(0) \) property will prevail when cointegration tests are applied, since samples for individual vintages vary. The HP filter was applied to the full available sample for each vintage.\(^{21}\)

All of the time series for \( R_t \) shown in Figure 2 are plotted in a continuous fashion until 1970. Thereafter only the last two years of data are shown for each one of the benchmark revisions considered. Once again, the purpose is to highlight the relatively larger values of \( R_t \) at the end of each sample. All of the series are found to be \( I(0) \) at conventional significance levels (see Table 1), with the exception of the revisions between the 2001 and 2004 vintages. Also notice that the longer the temporal difference between \( X_t^i \) and the particular \( X_t^{iV} \) examined, generally the larger are the revisions. The results are virtually identical when \( R_t \) is defined as the difference between vintages of real GDP growth (plot not shown) except that \( (X_t^{2004Q1} - X_t^{2001Q1}) \) is found to be \( I(1) \).\(^{22}\) The main point that bears repeating, however, is that the relatively larger revisions are a feature of the end of the sample, even when growth rates are used, and these already take account of the unit root property of the series considered.

**FIGURE 2 ABOUT HERE**

By way of introduction to the estimates and forecasts of Phillips curves presented below, we consider whether benchmark revisions and their behavior toward the end of the sample might reflect some underlying economic phenomenon that is only captured by subsequent benchmark revisions. Alternatively, it may be that the mere temporal distance between data revisions by itself gives a rough indication of the possibility of relatively larger revisions nearer the end of the sample. Table 2 gives

\(^{21}\) The data were also generated for cases where the HP filter was applied to \( X_t^i \) for the same sample as that available for each \( X_t^{iV} \) without affecting the conclusions described below. It is not immediately clear that this is the appropriate way to construct \( R_t \) since one might wish to retain potential differences in long-run trends between \( X_t^i \) and the 2004Q1 vintage to emphasize the role of these long-run factors in influencing the time-series properties of benchmark revisions.

\(^{22}\) The unit root tests produce decidedly more mixed results when growth in M1, M2, and real consumption spending are used in evaluating \( R_t \) (results not shown).
estimates of the simple correlation between pairs of revisions, again for estimates of the output gap.\(^{23}\) Taking the most recent revision as the benchmark (i.e., \(R_{2001Q1}^{2004Q1} = X_{2004Q1}^{2001Q1} - X_{2001Q1}^{2001Q1}\)), it is generally the case that neighboring revisions are highly correlated while the correlations drop rather quickly the longer the time span between revisions. Thus, for example, \(\rho(R_{2001Q1}^{2001Q1}, R_{1976Q1}^{1976Q1})\) is only 0.19, while \(\rho(R_{1976Q1}^{1976Q1}, R_{1971Q1}^{1971Q1})\) is 0.70. The extent to which these results reflect corrections or improvements made as a result of successive benchmark revisions is unclear. However, the correlations do suggest that distance in time across vintages is a rough indicator of the importance of such revisions, as well as an indication that the choice of vintage used to estimate some economic model could potentially significantly affect coefficient estimates. Further, the results in Table 2 also reveal that the simple correlations across revisions are almost always positive. Hence, differences between vintages at the time of benchmark revisions may have some predictive power for subsequent revisions.\(^{24}\)

Table 2 also displays correlations between equity and housing returns, and benchmark revisions.\(^{25}\) These are almost always negative, and in several instances, quite large. It is interesting, for example, that the correlations for revisions between 1976, 1981, 1986, 1991, and 2004 are highly negatively correlated with these asset returns. One possible explanation is that higher asset returns portend future structural changes in the economy which, if properly captured by the statistical agency, ought to be eventually incorporated into subsequent benchmark revisions.\(^{26}\)

TABLE 2 ABOUT HERE

\(^{23}\) To the extent that \(R_t \sim I(0)\), these simple correlations are informative about the linear relationship between revisions. The same would not be true of the log-levels of the series which, as discussed earlier, are generally found to be \(I(1)\).

\(^{24}\) Part of the explanation for these correlations is no doubt an overlapping data problem, but depending on the nature and magnitude of the benchmark revisions, this cannot be the entire story. The findings in Table 2 carry over to differences in vintages of real GDP growth.

\(^{25}\) Equity returns are the log difference in the Dow Jones Industrial Average, taken from the International Monetary Fund’s IFS CD-ROM (March 2006 edition). Housing price data were taken from the Office of Federal Housing Oversight House Price Index (for the U.S.) available at www.ofheo.gov/download.asp. Log differencing was also applied to this time series.

\(^{26}\) Another candidate series, albeit one that might also be indirectly captured by the other series considered here, is output per person hour, that is, a proxy for productivity growth. We did not consider this series; however, see van Norden (2005).
3.3. Cointegration and Cointegration Breakdown Tests

The unit root tests shown in Table 1 assume a particular form for the cointegrating relationship, namely that $\theta_0$ in (5) is unrestricted but $\theta_1$ is restricted to unity. We may instead wish to formally estimate and test the joint restriction that $\theta_0 = 0, \theta_1 = 1$, i.e., whether differences between vintages of data are indeed stationary. The results in Table 3 show that $X_t^v$ and $X_t^f$ are almost never cointegrated for the log of real GNP (output) or the log of real consumption expenditures (consumption). The results do not appear to be an artifact of the choice of lags in the Johansen testing procedure, nor were the test results especially sensitive to the treatment of the trend.

TABLE 3 ABOUT HERE

Results for the output gap, in contrast, suggest that differences between successive revisions and final revised data are independent random walks. As noted previously, by construction, the output gap is $I(0)$. However, because the HP filter is applied to the full available sample for each vintage, this implies that the stationarity property need not hold for the sub-sample over which the cointegration test is performed. The reason is the end-point problem associated with HP filtering. Whether this finding can be laid at the feet of the HP filter entirely is unclear. In any event, while the random walk property permits the researcher to carry on with estimation, there may still be useful information that is lurking in the data that may be worth modeling, such as whether structural factors that could have produced large, but temporary, departures from trend output.

We now turn to the cointegration breakdown tests. The results are plotted in Figure 3, where the top and bottom portions of the figure show $p$-values for the Andrews and Kim (2006) $P_c$ and $R_c$ statistics, respectively, corresponding to a test of the null hypothesis of stability at the end of each
The vertical lines indicate the year when a benchmark revision took place (i.e., every five years beginning in 1966 and ending in 2001). The \( p \)-values are shown for the last three years prior to a benchmark revision.

**FIGURE 3 ABOUT HERE**

There are a number of instances where the \( p \)-values are below 0.05 or 0.10 at the end of the sample for the different vintages in both panels of Figure 3, indicating rejection of the null hypothesis of stability. In general, the results point to the presence of cointegration breakdown toward the end of each sample when vintages are compared across benchmark revisions.

### 4. Benchmark Revisions and Inflation Forecasts

We now consider whether benchmark revisions incorporate useful information that could have been used to improve forecasts of inflation. Both private and public sector forecasts are considered as well as forecasts generated from standard Phillips Curve specifications. In the case of private sector forecasts, we simply ask whether, in addition to last period’s inflation rate differences, benchmark revisions and final estimates might have improved forecasts. In this case, the specification is given by

\[
\pi_t \text{for} = \alpha_0 + \alpha_1 \pi_{t-1} + \alpha_2 u_{t-1} + \epsilon_t, \tag{9}
\]

where \( \pi_t \text{for} \) is the inflation forecast released by a public or private sector agency (see Section 3.1), \( \pi_{t-1} \) is the actual lagged inflation rate, and \( u_{t-1} \) is the error-correction term from (5), that is, the lagged difference between a particular vintage of real GDP and the final revision.\(^{28}\) Equation (9) assumes that the current

\(^{27}\) The previous section briefly describes how the \( p \)-value can be computed using sub-sampling. Detailed definitions and formulas for the test statistics can be found in Andrews and Kim (2006).

\(^{28}\) In what follows we only consider revisions to real GDP. A few tests with real personal consumption expenditures, M1, and M2 yielded qualitatively similar conclusions. One possibility ignored here is that the appropriate residual is one that takes into account a break somewhere near the end of the sample. One difficulty is that there may be more than one candidate for a break. Second, the view taken here is that the relevant information comes from the error-correction term, uncorrected for a break. The problem is not that there is no cointegration but that cointegration
forecast for inflation at time $t$ is related to last period’s realized inflation as well as possibly some linear combination of two successive benchmark revisions. The superscript indicates the vintage used. Our interest is to test whether $\alpha_2$ is significantly different from zero.\footnote{The $u_t$ in (8) refers to revisions between a particular benchmark revision and the latest available data vintage (2004Q1). Clearly, there are other pairs of revisions that could be considered (e.g., between data, say, in 2001 and an earlier benchmark revision), and it is also conceivable that some combination (possibly linear or non-linear) between a series of benchmark revisions (overlapping or not) could also be used in place of $u_{t-1}$. These extensions are not considered, as the principal aim of this paper is only to demonstrate the potential for some benchmark revision to improve inflation forecasts.} If benchmark revisions contain information that could have improved inflation forecasts, then we would expect rejection of the null hypothesis that $\alpha_2 = 0$.

Instead we can ask, in the context of some model of inflation such as a Phillips curve, whether forecasts from such a model could have been improved using information contained in differences between benchmark revisions and final revised estimates, as defined in this paper. Our estimated forward-looking Phillips curve is of the form,

$$\pi_t^f = \beta_0 + \beta_1 \tilde{y}_t^v + \beta_2 \pi_t^f + \beta_3 E(\pi_{t-1}^f | \pi_{t-1}^f) + \beta_4 u_{t-1}^v + \xi_t^v,$$

(10)

where $\tilde{y}_t^v$ is an estimate of the output gap, $E(\pi_{t-1}^f | \pi_{t-1}^f)$ is the expectation of next period’s inflation rate ($\pi_{t-1}^f$) conditional on information up to lagged inflation ($\pi_{t-1}^f$), and all other terms have previously been defined. Once again, the superscript indicates the vintage used. Equation (10) is a widely used specification for a Phillips curve that combines both forward and backward-looking elements. In addition to (10), two other variants are estimated. In one version we restrict $\beta_3 = 0$ so that the Phillips curve is of the backward-looking variety. In another, case we estimate a version of (10) with $\beta_2 = 0$ so that the resulting specification is purely forward-looking as far as the role of inflation is concerned. A forward-looking Phillips curve model cannot be properly estimated via OLS when $\beta_3 \neq 0$, unlike the backward-looking Phillips curve which can be estimated via OLS. When the two variants that incorporate forward-
looking elements are estimated, we rely (as does much of the extant literature) on the Generalized Method of Moments (GMM) approach. In relying on GMM, the choice of instruments can be crucial, and there is an emerging literature that is beginning to address an old problem from a variety of new perspectives.\(^{30}\)

In what follows, first, the usual, but not always ideal, strategy of using lagged values of \(\bar{y}_t\) and \(\pi_t\) as instruments is adopted. Next, three lags in benchmark revisions are added to determine whether this improves our estimates of forward-looking Phillips curves. Improvements in estimates are judged not only according to Hansen’s \(J\)-test for over-identifying restrictions, the common metric for reporting whether the chosen instruments are adequate, but we also add Stock’s \(F\)-test for instrument adequacy and Andrews’ GMM information criterion (GMMC) for instrument relevance (see Hall, 2005). Space limitations prevent going into detail concerning the usefulness of these tests. However, readers are referred to Hall (2005) for an extensive discussion of these tests.

Table 4 presents the results. To conserve space, we focus on two vintages of data, namely, the 1996 and 2001 benchmark revisions. There is some evidence that the error correction-term based on the benchmark revisions could have improved all types of forecasts for at least one of the two benchmark revisions shown. Turning to estimates of various Phillips curve specifications, the evidence on the information content of the error-correction terms is mixed. When backward-looking models are estimated, the error-correction term is significant for the 1996 benchmark revision but not the 2001 benchmark revision. In the case of forward-looking models, only the 2001 vintage revision series is statistically significant when the models are augmented with the error-correction term. When benchmark revisions are added to the list of instruments in GMM estimation, there is some evidence that this improves Phillips curve estimates for the 1996 benchmarks based on the GMMC and \(F\)-test statistics, but there is no appreciable improvement for the 2001 benchmark. While the evidence is not overwhelming, there is some evidence that data revisions can improve Phillips curve estimates and be potentially useful in a forecasting exercise. At the very least, the potential for such revisions to matter ought to be recognized.

\(^{30}\) See, e.g., the references contained on James Stock’s “Weak Instruments Web Page” available at; http://ksghome.harvard.edu/~JStock/aus/websupp/index.htm.
While the potential usefulness of benchmark revisions for inflation forecasts in-sample does suggest that such revisions contain useful information, a stricter test of the information content in such revisions requires that we consider out-of-sample forecasts. Part (a) of Figure 4 plots U.S. inflation for the period 1960 to 2004. The three vertical long dashed lines identify three benchmark years, namely, 1981, 1991, 1996, and 2001. These benchmark years will be the subject of the out-of-sample inflation forecasts to be reported below. The short vertical dashed lines represent the end-points of the various samples over which Phillips curves are estimated. Only the forward-looking Phillips curve is considered for this experiment.\(^{31}\) The gap between the two vertical dashed lines highlights the out-of-sample period over which inflation forecasts are generated. In Table 5, we report both the one-step-ahead forecast, together with its standard error, as well as the RMSE for one-year-ahead (i.e., four-step-ahead) forecasts. The forecast samples were chosen in order to assess the sensitivity of the out-of-sample forecasts to the level of inflation, and whether the U.S. economy was in a recession. The vertical shaded areas highlight recessionary periods as identified by NBER business cycle reference dates (http://www.nber.org/cycles.html/). Inflation is measured as 100 times the fourth-order log change in the CPI, using the real-time data available in August 2006 from the Federal Reserve Bank of Philadelphia’s real-time data set.

Part (b) of Figure 4 plots one- through four-step-ahead forecasts of inflation for the various cases considered. The forward-looking Phillips curve FL1, shown in Table 4, is estimated with or without the inclusion of benchmark revisions for a sample that always begins in 1960Q1 and ends in 1994Q4 in the

\(^{31}\) More precisely, only version FL1 among the forward-looking Phillips curve reported in Table 4 is considered.
case of the 1996 vintage. For the 1981 vintage, the estimation period ends in 1979Q4; for the 1991 vintage, the estimation period ends in 1989Q4; for the 2001 vintage the sample is 1960Q1–1999Q4. Subject to data limitations, inflation forecasts are generated out-of-sample for all available vintages. Thus, for example, as shown in part (b) of Figure 4, forecasts for the year 1990 are generated by estimating a forward-looking Phillips curve using vintages 1991, 1996, and 2001. Similarly, vintages 1996, and 2001 are used to forecast out-of-sample for 1995, while vintages 1981, 1996, and 2001, are used to forecast out-of-sample for 1980. Only one out-of-sample forecast for 2000 is generated, namely relying on 2001 vintage data. Clearly, other vintages and combinations of benchmark revisions could have been selected, but the out-of-sample forecasts generated cover a wide range of inflationary and recessionary experiences in the U.S. economy. The results in Table 5 clearly show that the RMSE for out-of-sample forecasts that omit the benchmark revisions are higher than when these revisions are included in the specification. Moreover, it is usually the case that the standard error of the one-step-ahead forecast is also often considerably higher when the error-correction term is omitted from the forecasting model. Lastly, while the specification that includes the impact of benchmark revisions generally underestimates inflation one-year-ahead, the opposite is true for the specification which omits these same lagged revisions.

FIGURE 4 ABOUT HERE

5. Conclusions

This paper considers the time-series properties of benchmark revisions to key U.S. macroeconomic aggregates such as real GDP. Evidence is presented to the effect that benchmark revisions are not cointegrated because there is a breakdown of the cointegration property toward the end of the sample. Hence, we conclude that there is potentially some information content in benchmark revisions. Estimates of Phillips curves augmented with benchmark revisions treated as an error-correction term produce improvements in inflation forecasts. A number of extensions and other considerations remain to be dealt with in future research. First, it may be useful to explore whether some combination of
benchmark revisions could help improve inflation forecasts or whether existing forecasts or forward-looking models of inflation anticipate the information content of benchmark revisions. Second, even if there is useful information content in benchmark revisions, most, though not all, of the various estimated specifications rely on a generated regressor which may raise additional econometric issues. Third, this paper only investigates the impact of benchmark revisions at more or less regularly spaced intervals of time. It may be that revisions incorporate useful information at more frequent intervals than relied on in this paper. Fourth, unless we are able to specify a working model of benchmark revisions we will not be able to identify whether, say, changes in definitions, re-basing, or some other factor can explain the potential role of revisions in forecasting inflation. Fifth, the results indicate that monetary policy in a “data rich” environment (Bernanke and Boivin, 2003) might also consider a role for the impact of revisions. Lastly, only end-of-sample cointegration breakdowns are considered. Clearly, cointegration could fail earlier in the data considered, perhaps even at the beginning of each sample. These, and other extensions, are left for future research.
Acknowledgements

An earlier version of this paper was presented at the workshop of Macroeconomic Forecasting, Analysis, and Policy with Data Revisions at CIRANO, Montreal, October 2006, the 27th International Symposium on Forecasting in New York, and the Far Eastern Meeting of Econometric Society in Taipei, Taiwan. Comments by Øyvind Eitrheim, Mark Wohar, and workshop and conference participants are gratefully acknowledged.
References


Figure 1: Stylized facts about benchmark revisions, logarithm of real GDP

Notes: The top panel is for all vintages and observations; the bottom panel is for all vintages, last two years of data. See the text for data sources and series description.
Figure 2: Benchmark revisions in the output gap

Notes: See the text. Last two years of data only shown for all vintages after 1970. Also see Figure 1. The output gap is computed via HP filtering (see text for estimation details) applied to the full available sample to each series individually.
Figure 3: End of sample cointegration breakdown tests

A. $P_C$ statistic

![Graph showing $P_C$ statistic and critical value]

B. $R_C$ statistic

![Graph showing $R_C$ statistic and critical value]

Note: See (5) for the cointegrating test equation. The vertical bars identify the benchmark revisions. For example, the bar labeled 1966, and the values to the left of this vertical bar show the test statistic and critical values for the case where $v = 1966$ and $f = 2004$. 
Figure 4: U.S. inflation and out-of-sample inflation forecasts

A. Full sample

Note: Vertical bars in the top figure indicate the out-of-sample forecasting period which is a function of the vintage used. The construction of the inflation series is described in the text. The bottom figure plots actual inflation and one-step-ahead forecasts of inflation for a four-quarter horizon. The legend (e.g., 2001V80) indicates the vintage year used (e.g., 2001) and the end of the estimation sample (here, 1980:4).
Table 1: Unit root tests for benchmark revisions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Real consumption</td>
<td>1.13 (4)</td>
<td>0.42 (4)</td>
<td>0.01 (1)</td>
<td>-0.18 (0)</td>
<td>-0.40 (2)</td>
<td>-0.77 (2)</td>
<td>-1.95 (1)</td>
<td>-0.34 (2)</td>
</tr>
<tr>
<td></td>
<td>[0.98, 71]</td>
<td>[0.98, 91]</td>
<td>[0.96, 115]</td>
<td>[0.94, 135]</td>
<td>[0.90, 153]</td>
<td>[0.83, 173]</td>
<td>[0.31, 143]</td>
<td>[0.92, 213]</td>
</tr>
<tr>
<td>Real GDP</td>
<td>0.70 (1)</td>
<td>0.16 (3)</td>
<td>0.92 (4)</td>
<td>-0.38 (5)</td>
<td>1.44 (0)</td>
<td>0.62 (0)</td>
<td>3.70 (2)</td>
<td>-1.46 (6)</td>
</tr>
<tr>
<td></td>
<td>[0.99, 74]</td>
<td>[0.97, 92]</td>
<td>[0.99, 111]</td>
<td>[0.91, 130]</td>
<td>[0.99, 155]</td>
<td>[0.99, 175]</td>
<td>[1.00, 142]</td>
<td>[0.55, 209]</td>
</tr>
<tr>
<td>Output Gap</td>
<td>-4.55 (0)</td>
<td>-6.98 (0)</td>
<td>-5.20 (0)</td>
<td>-3.07 (4)</td>
<td>-5.84 (0)</td>
<td>-4.96 (0)</td>
<td>-5.13 (0)</td>
<td>-0.67 (4)</td>
</tr>
<tr>
<td></td>
<td>[0.00, 75]</td>
<td>[0.00, 95]</td>
<td>[0.00, 115]</td>
<td>[0.03, 131]</td>
<td>[0.00, 155]</td>
<td>[0.00, 175]</td>
<td>[0.00, 144]</td>
<td>[0.85, 211]</td>
</tr>
<tr>
<td>∆log(real consumption)</td>
<td>-6.01 (2)</td>
<td>-3.35 (4)</td>
<td>-4.51 (7)</td>
<td>-2.46 (4)</td>
<td>-2.13 (8)</td>
<td>-5.47 (4)</td>
<td>-1.61 (4)</td>
<td>-5.71 (11)</td>
</tr>
<tr>
<td></td>
<td>[0.00, 69]</td>
<td>[0.02, 87]</td>
<td>[0.00, 104]</td>
<td>[0.13, 127]</td>
<td>[0.23, 143]</td>
<td>[0.00, 167]</td>
<td>[0.48, 136]</td>
<td>[0.00, 200]</td>
</tr>
<tr>
<td>∆log(real GDP)</td>
<td>-2.94 (5)</td>
<td>-3.19 (5)</td>
<td>-4.07 (1)</td>
<td>-4.40 (0)</td>
<td>-3.83 (4)</td>
<td>-4.46 (4)</td>
<td>-4.53 (0)</td>
<td>-3.66 (4)</td>
</tr>
<tr>
<td></td>
<td>[0.05, 66]</td>
<td>[0.02, 86]</td>
<td>[0.00, 110]</td>
<td>[0.00, 131]</td>
<td>[0.00, 147]</td>
<td>[0.00, 167]</td>
<td>[0.00, 140]</td>
<td>[0.01, 207]</td>
</tr>
</tbody>
</table>

Notes: All tests are based on the Augmented Dickey-Fuller test statistic with the lag augmentation chosen according to the AIC. No deterministic trends are allowed in any test equation. The lag length in the augmentation is given in parentheses. The p-value corresponding to the null hypothesis of a unit root followed by the number of observations after differencing are given in square brackets. For the 1996 and 2001 vintages, samples begin in 1959Q1; otherwise, they begin in 1947Q1. All samples end with the fourth quarter of the year prior to the benchmark revision (e.g., 1965Q4 for the 1966 vintage). All time series tested are \( R_t \) as defined in (2), where \( X^v_t \) is the 2004 vintage, and \( X^v_t \) are for \( v = 1996, \ldots, 2001 \). Bold items signify rejections of the null of a unit root in \( R_t \) at conventional significance levels.
Table 2: Simple correlations between benchmark revisions, output gaps, and asset prices

<table>
<thead>
<tr>
<th>Benchmark revisions</th>
<th>Asset prices</th>
<th>Benchmark revision year</th>
<th>Equities (S)</th>
<th>Housing (H)</th>
<th>Fed funds (ΔFFR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5) (6) (7)</td>
<td>(8)</td>
<td>(9)</td>
<td>(10)</td>
<td>(11)</td>
</tr>
<tr>
<td>1966</td>
<td>0.97</td>
<td>1966</td>
<td>0.02</td>
<td>-</td>
<td>-0.49</td>
</tr>
<tr>
<td>1971</td>
<td>0.70</td>
<td>1971</td>
<td>-0.20</td>
<td>-</td>
<td>-0.47</td>
</tr>
<tr>
<td>1976</td>
<td>0.91</td>
<td>1976</td>
<td>-0.53</td>
<td>-</td>
<td>-0.46</td>
</tr>
<tr>
<td>1981</td>
<td>0.66</td>
<td>1981</td>
<td>-0.33</td>
<td>-0.02</td>
<td>-0.30</td>
</tr>
<tr>
<td>1986</td>
<td>0.98</td>
<td>1986</td>
<td>-0.05</td>
<td>-0.39</td>
<td>-0.16</td>
</tr>
<tr>
<td>1991</td>
<td>0.31</td>
<td>1991</td>
<td>0.00</td>
<td>-0.34</td>
<td>-0.16</td>
</tr>
<tr>
<td>1996</td>
<td>-0.06</td>
<td>1996</td>
<td>0.02</td>
<td>-0.12</td>
<td>-0.25</td>
</tr>
<tr>
<td>2001</td>
<td>-0.04</td>
<td>2001</td>
<td>-0.93</td>
<td>-0.15</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

Note: Columns (1)–(7) report the simple correlation between pairs of benchmark revisions (e.g., the correlation between $R_{1966}^t$ and $R_{1981}^t$ is 0.68). Columns (9)–(11) report the simple correlations between equity returns (S), housing price inflation (H), the change in the fed funds rate (ΔFFR) and the particular benchmark revision listed in column (8) (e.g., the simple correlation between $R_{1991}^t$ and H is -0.34).
Table 3: Johansen VAR-based cointegration test Results, vintage pairs from 1966 to 2004

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>(2) $r = 0$ 5.86</td>
<td>(1) $r = 0$ 4.58</td>
<td>(2) $r = 0$ 49.34*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 2.03</td>
<td>$r = 1$ 0.02</td>
<td>$r = 1$ 0.35*</td>
</tr>
<tr>
<td>1996</td>
<td>(2) $r = 0$ 14.10</td>
<td>(1) $r = 0$ 26.97*</td>
<td>(1) $r = 0$ 20.20*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 1.34</td>
<td>$r = 1$ 1.34</td>
<td>$r = 1$ 12.33*</td>
</tr>
<tr>
<td>1991</td>
<td>(2) $r = 0$ 12.13</td>
<td>(1) $r = 0$ 3.33</td>
<td>(2) $r = 0$ 44.23*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 0.18</td>
<td>$r = 1$ 1.89</td>
<td>$r = 1$ 20.97*</td>
</tr>
<tr>
<td>1986</td>
<td>(2) $r = 0$ 9.29</td>
<td>(1) $r = 0$ 2.93</td>
<td>(2) $r = 0$ 42.07*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 0.12</td>
<td>$r = 1$ 0.05</td>
<td>$r = 1$ 29.62*</td>
</tr>
<tr>
<td>1981</td>
<td>(2) $r = 0$ 12.72</td>
<td>(1) $r = 0$ 11.26</td>
<td>(2) $r = 0$ 35.47*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 0.87</td>
<td>$r = 1$ 2.87</td>
<td>$r = 1$ 10.50*</td>
</tr>
<tr>
<td>1976</td>
<td>(2) $r = 0$ 14.26</td>
<td>(1) $r = 0$ 5.65</td>
<td>(2) $r = 0$ 35.98*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 3.84</td>
<td>$r = 1$ 0.07</td>
<td>$r = 1$ 18.55*</td>
</tr>
<tr>
<td>1971</td>
<td>(2) $r = 0$ 12.85</td>
<td>(1) $r = 0$ 4.61</td>
<td>(2) $r = 0$ 34.67*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 0.27</td>
<td>$r = 1$ 0.05</td>
<td>$r = 1$ 24.50*</td>
</tr>
<tr>
<td>1966</td>
<td>(2) $r = 0$ 9.77</td>
<td>(1) $r = 0$ 5.65</td>
<td>(2) $r = 0$ 24.48*</td>
</tr>
<tr>
<td></td>
<td>$r = 1$ 0.75</td>
<td>$r = 1$ 1.56</td>
<td>$r = 1$ 14.10*</td>
</tr>
</tbody>
</table>

Notes: Lag lengths are shown in parenthesis and chosen according to the SIC. The series are assumed to contain a constant and a linear trend, and the form of the cointegrating equation is as in (5). $r = 0$ ($r = 1$) is the maximal eigenvalue test statistics for the null that there is no cointegration (at most one cointegrating vector); * indicates significant at the 5% level.
Table 4: Inflation forecasts and the significance of benchmark revisions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.20</td>
<td>1.81</td>
<td>0.92</td>
<td>0.87</td>
<td>0.88</td>
<td>0.68</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(1.28)†</td>
<td>(0.16)**</td>
<td>(0.30)**</td>
<td>(0.17)**</td>
<td>(0.16)**</td>
<td>(0.14)**</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.02)**</td>
<td>(0.02)*</td>
<td>(0.02)**</td>
<td>(0.02)</td>
</tr>
<tr>
<td>$\pi_{t-1}$</td>
<td>0.47</td>
<td>0.56</td>
<td>0.89</td>
<td>0.90</td>
<td>0.78</td>
<td>0.80</td>
<td>0.99</td>
<td>0.99</td>
<td>0.45</td>
<td>0.46</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.09)**</td>
<td>(0.06)**</td>
<td>(0.10)**</td>
<td>(0.07)**</td>
<td>(0.03)**</td>
<td>(0.03)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
</tr>
<tr>
<td>$E(\pi_{t-1}</td>
<td>\pi_{t-1})$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.56</td>
<td>0.56</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.03)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
<td>(0.02)**</td>
</tr>
<tr>
<td>$\bar{y}_{t-1}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.19</td>
<td>0.19</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.02)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
<td>(0.01)**</td>
</tr>
<tr>
<td>$\nu_{t-1}$</td>
<td>-0.037</td>
<td>-0.32</td>
<td>-0.53</td>
<td>-0.01</td>
<td>-0.80</td>
<td>-0.44</td>
<td>0.31</td>
<td>0.08</td>
<td>-0.09</td>
<td>-0.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.24)†</td>
<td>(0.11)**</td>
<td>(0.25)*</td>
<td>(0.11)†</td>
<td>(0.42)†</td>
<td>(0.33)†</td>
<td>(0.14)**</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.20)**</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>0.52</th>
<th>0.73</th>
<th>0.79</th>
<th>0.82</th>
<th>0.82</th>
<th>0.83</th>
<th>0.98</th>
<th>0.99</th>
<th>0.99</th>
<th>0.99</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>[0.79]</td>
<td>[0.85]</td>
<td>[0.87]</td>
</tr>
<tr>
<td>GMMIC</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-24.73</td>
<td>-25.42</td>
<td>-29.66</td>
</tr>
<tr>
<td>$F$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>718.1</td>
<td>891.9</td>
<td>793.2</td>
<td>629.4</td>
<td></td>
</tr>
</tbody>
</table>

Note: Variables in the first column are defined in the main body of the paper. CONS = Consensus forecasts. ECON = forecasts from The Economist. OECD = forecasts from the OECD Main Economic Outlook. BL = backward-looking Phillips curve. FL = forward-looking Phillips curve. BL specifications are estimated via OLS. FL specifications are estimated via GMM. $J$ is Hansen’s test for over-identifying restrictions (p-values for the Hansen test statistic are given in square brackets). GMMIC is the Andrew’s GMM Information Criterion to Andrews. $F$ is the Stock’s test for instrument adequacy. Standard errors are reported in parenthesis. †,*,** indicate significance at the 10%, 5%, and 1% levels, respectively.
<table>
<thead>
<tr>
<th>Vintage</th>
<th>Estimation Period</th>
<th>RMSE with $u_{t-1}$</th>
<th>RMSE without $u_{t-1}$</th>
<th>One-step-ahead forecast (s.e.) with $u_{t-1}$</th>
<th>One step-ahead forecast (s.e.) without $u_{t-1}$</th>
<th>Actual inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>1960.1-1994.4</td>
<td>0.47</td>
<td>0.58</td>
<td>3.10% (0.33)</td>
<td>2.16 (0.13)</td>
<td>2.64%</td>
</tr>
<tr>
<td>1996</td>
<td>1960.1-1989.4</td>
<td>0.90</td>
<td>1.11</td>
<td>4.27% (0.31)</td>
<td>3.85 (0.13)</td>
<td>4.81%</td>
</tr>
<tr>
<td>1996</td>
<td>1960.1-1979.4</td>
<td>1.06</td>
<td>1.61</td>
<td>10.89% (0.11)</td>
<td>10.88 (0.11)</td>
<td>11.62%</td>
</tr>
<tr>
<td>2001</td>
<td>1960.1-1999.4</td>
<td>0.66</td>
<td>0.78</td>
<td>3.85% (0.12)</td>
<td>3.73 (0.13)</td>
<td>2.55%</td>
</tr>
<tr>
<td>2001</td>
<td>1960.1-1994.4</td>
<td>0.57</td>
<td>0.68</td>
<td>3.26% (0.24)</td>
<td>2.12 (0.12)</td>
<td>2.64%</td>
</tr>
<tr>
<td>2001</td>
<td>1960.1-1989.4</td>
<td>0.71</td>
<td>1.16</td>
<td>3.99% (0.09)</td>
<td>4.21 (0.25)</td>
<td>4.81%</td>
</tr>
<tr>
<td>2001</td>
<td>1960.1-1979.4</td>
<td>1.07</td>
<td>1.61</td>
<td>10.88% (0.12)</td>
<td>10.76 (0.13)</td>
<td>11.62%</td>
</tr>
</tbody>
</table>

Notes: RMSE= root mean squared forecast error. Estimates are based on equation FL1 shown in Table 4. Each model was estimated and dynamic forecasts were generated four quarters ahead, and the table reports RMSEs for those forecasts. The one-step-ahead forecast is for the quarter that immediately follows the end of the estimation sample.