WHAT CAN WE LEARN FROM COMPREHENSIVE DATA REVISIONS FOR FORECASTING INFLATION? SOME U.S. EVIDENCE

Pierre L. Siklos
Department of Economics
Wilfrid Laurier University

and

Viessman Research Centre
Waterloo, ON
Canada

[Preliminary draft: April 2006]
[First complete draft: June 2006]
[This draft: October 2006]

"What Can We Learn from Data Revisions for Forecasting Inflation? Some U.S. Evidence" in M. Wohar and D. Rapach (2008) **Forecasting in the Presence of Structural Breaks and Model Uncertainty**[Frontiers of Economics and Globalization Series, VOL. 3] (Amsterdam: Elsevier/Emerald), pp. 271-299.

Abstract

This paper examines the empirical properties of benchmark revisions to key U.S. macroeconomic aggregates. The news versus noise impact of revisions is interpreted in the context of the cointegration property of successive benchmark revisions. It is found that the cointegration property breaks down in the last year or so 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.

Pierre L. Siklos, Department of Economics and Viessmann European Research Centre

e-mail: psiklos@wlu.ca

Home Page: http://www.wlu.ca/sbe/psiklos

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 what was believed to have been 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 (e.g., also see Runkle (1998), and 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) update the data in light of the evolution of the U.S. economy, reflect changes in definitions for some series, and include newly available or

[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, and www.philadelphiafed.org/econ/forecast/reaindex.html for updates to US real time series. The movement to develop real-time data sets has become an international phenomenon. See, for example, 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/.

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 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 believed 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, for example, 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.[3] It is not immediately clear that one approach offers an advantage over another. Comparing revisions

[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).
[3] Implicitly, we consider only the case where the levels of each series is integrated of order one, or I(1), so that first differencing is all that is necessary to render such series I(0).

in growth rates can mitigate some of the minor changes in statistical or reporting procedures that could contaminate the levels data. 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 revision process.[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 and Kim (2003), and Andrews (2003), 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

[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.

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.[5] In any event, the economic mechanism that leads to the cointegration breakdown property needs to be spelled out. 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, our findings suggest that 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 curve type 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 three briefly describes the data, and presents some stylized facts, 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

[5] Another potentially important source of the breakdown in cointegration could be changes in trend productivity growth. This possibility is not investigated here. See, however, van Norden (2005).

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.[6] As a result, earlier vintages are optimal forecasts of subsequent revisions of the data. More formally, we can express the relationship between one vintage (*v*) and a subsequent one, which we refer to the "final" vintage (*f*), as follows:

$$X_t^v = X_t^f + \varepsilon_t \tag{1}$$

where X_t^v is a time series for, say, GDP, released temporally prior to the final release of the same series, denoted X_t^f . In this setup, and denoting $\rho(X)$ as the correlation coefficient, if

$$\rho(X_t^f, \varepsilon_t) = 0$$

then revisions are pure noise, since final revised data are uncorrelated with the error term, while, in the event that

$$\rho(X_t^v, \varepsilon_t) = 0$$

holds, 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 co-exist, are also possible and empirically more likely. The foregoing dichotomy is the one introduced by Mankiw, Runkle, and Shapiro (1984), and used by many others since with some modifications (e.g., Faust, Rogers, and Wright (2005)). Moreover, since we have not been specific about how temporally close or distant

[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.

X_t^v is from X_t^f, there is also the potential for $\rho(X)$ to be a function of ($v - f$). If we define revisions employing the following expression:

$$R_t \equiv X_t^v - X_t^f \tag{2}$$

then a regression of the form

$$R_t = \alpha + \beta X_t^v + u_t \tag{3}$$

, also known as the Mincer-Zarnowitz regression, is used to test efficiency which requires non-rejection of the null

$$H_0 : \alpha = \beta = 0$$

Since the focus of this paper in on the information content of benchmark revisions, a few words concerning the underlying statistical model of benchmark revisions is in order. Croushore and Stark (2006) specify the following statistical model for benchmark revisions

$$X_t^{v+s} = X_t^* + \eta_t^{v+s} \tag{4}$$

where $v+s$ indicates that we are only concerned with revisions s periods apart; η_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 the specification (4) may be correct, there is little compelling reason to justify it, except perhaps that this model is consistent with the 'surprising' result reported in Croushore and Stark (2002). Alternatively, one could argue that successive benchmarks correct past errors as well as improving 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

[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.

revisions. Therefore, the paper posits that the true value of X_t, namely X_t^*, 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^v and X_t^f potentially exists via the cointegrating one. Under this view, R_t is a stationary process, 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

$$X_t^v = \theta_0 + \theta_1 X_t^f + u_t \qquad (5)$$

As in the Mincer-Zarnowitz formulation, non-rejection of the null $\theta_0 = 0, \theta_1 = 1$ reduces (5) to equation (1). The residuals 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

$\theta_0 = 0, \theta_1 = 1$ 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 X_t^f, and earlier benchmark revisions as $X_t^{v_i}$, 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 VAR format as

$$\begin{pmatrix} X_t^f \\ X_t^{v_1} \\ \vdots \\ X_t^{v_i} \end{pmatrix} = \Phi x_t + \begin{pmatrix} \varepsilon_t^f \\ \varepsilon_t^1 \\ \vdots \\ \varepsilon_t^{v_i} \end{pmatrix} \tag{6}$$

where $X_t^{v_1},...,X_t^{v_i}$ are the different benchmark vintage that temporally precede X_t^f, and x_t is a vector containing lags of $X_t^{v_1},...,X_t^{v_i}$, X_t^f.

If there is cointegration this will be reflected in the rank of π in the error correction form

$$\Delta x_t^v = \pi x_{t-1}^v + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i}^v + \varepsilon_t \tag{7}$$

where all the terms were previous defined. In what follows, while (6) and (7) represent the most general form of the potential cointegrating relationship between successive vintages$_t$, we focus on the pair of vintages ($X_t^{v_i}$ 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 greatest for more recent observations than for more temporally distant observations. Hence, the breakdown of the cointegration property, if it occurs at all, is arguably greater 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 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 and Kim (2003), and Andrews (2003), are the appropriate ones under the present circumstances, since they can be applied to the end of a particular sample.[8]

[8] As we shall see, while the Andrews and Kim (2003) test solves one problem it leaves a question unanswered as it is a residuals-based test. When a test based on equation (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

Assume that a sample of data is split at $t = m$, where m 0[1, T] and T is the number of observations in the sample. If we write the cointegrating relationship as in equation (5), and there is a breakdown in cointegration in the sub-sample [m, T], then the linear relationship between X_t^v and X_t^f can be written

$$y_t = \begin{cases} x_t' \theta_0 + u_t, t = 1,...,T \\ x_t' \theta_t + u_t, t = T+1,....,T+m \end{cases} \quad (8)$$

The regressor $x_t' (= X_t^f))$ is the case depicted in equation (5) with $y_t = X_t^v$, but other variables can be included whether they are $I(1)$, stationary, or incorporate a deterministic component. If the vintages are cointegrated, and their relationship is stable, then this is akin to testing the null

$$H_0: \theta_0 = \theta_t \ \forall_t = T+1,...,T+m$$

in which case $u_t \sim I(0)$. The alternative is $\theta_0 \neq \theta_t$, for some $t \in \{T+1,...,T+m\}$. As explained in Andrews and Kim (2003), the breakdown of cointegration can occur either because θ changes or because u_t ceases to be I(0), as would be true if the cointegration property holds, and becomes I(1). The relevant tests are of the Chow variety, and the choice of test statistics is sensitive to the location of the split in the sample. Andrews and Kim (2003) show that, given the empirical distribution function of $\sum_{t=j}^{j+m-1} u_t^2,\ j = 1,...,T-m+1$, the computation of critical values is straightforward. A second set of tests is based on $\sum_{i=T+1}^{T+m} (\sum_{j=T+1}^{T+m} \hat{u}_j)^2$. Andrews and Kim (2003) conduct an extensive series of power and size tests, and conclude that the tests based on the

---

was a failure in a previous pair of vintages, then subsequent revisions, and forecasts, will have made the relevant adjustments. In any event, in what follows, we will not consider these complications or extensions.

foregoing test statistics are recommended.[9] Critical values are determined via parametric sub-sampling and not on asymptotic theory.[10]

To summarize, we would normally expect u_t in equation (5) to at least be stationary, that is, $u_t \sim I(0)$.[11] Recall that, for the purposes of the exercise conducted here, u_t represents the residual 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 in hindsight appears to be important benchmark revisions that portend significant structural changes in the US economy. As a result, certain benchmark revisions at least 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 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.

---

[9] If θ in equation (8) is estimated over the sample t=1, …, T we denote it $\hat{\theta}_{1-T}$. Andrews and Kim (2003) 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.

[10] Essentially, this means that the distribution of the test statistic is derived from sequentially applying the test to a stable sample. See Andrews (2003).

[11] Since the metric is, in part, whether u_t is 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.

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 2001). The present paper focuses on the eight benchmark revisions. They are: 1966, 1971, 1976, 1981, 1986, 1992, 1996, and 2001, all occurring in February of those years. Hence, we take these revisions as having occurred in the first quarter of the benchmark revision year. The vintage 2004Q1 (or, more precisely, February 2004) represents the "final" vintage in our data set (i.e., proxies X_t^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.[12]

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 H-P filter (with smoothing parameter 1600) to the log level of real GDP. Standard H-P detrending has well-known drawbacks, such as the end-point problem, and the possibility of inducing spurious cycles, but most researchers employ this

[12] 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 whether, for the purposes of this paper, base year changes should be treated differently from other definitional changes. There are a variety of reasons for such changes, and not all reflect purely rebasing. It should also be noted that rebasing is also done from time to time to account for changing patterns in consumption spending by households.

filter to create a proxy for the output gap.[13] 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 (e.g., Kozicki (2004), and references therein).[14] 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 are: 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.[15] 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, while Consensus and Economist forecasts only begin in 1990, while whereas SPF data begin in 1981.[16] 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

[13] Some experiments that rely on the measure of potential output data from the Congressional Budget Office, a widely-used time series did not change any of our conclusions (www.cbo.gov).
[14] Again some experimentation with different vintages of the CPI confirms this to be the case.
[15] 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 highest frequency.
[16] 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.

forecasts could have been improved by the addition of benchmark revisions to real GDP in the estimated specifications.[17]

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, overall, 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, based at least 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), and several other authors since, have reported revisions in growth rates are analyzed instead of ones in the levels. Second, visually there are often large revisions near the end of each sample for most of the pairs shown in the figure. Third, 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 show 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

[17] Also considered was the impact of revisions in real personal 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.

confirmed by the data.[18] This will become even more apparent, as we shall see, when we examine first differences in the log of real GDP.

Figure 2 plots several estimates of R_t for the output gap but where X_t^f is an estimate of the output gap for the 2004Q1 vintage while $X_t^{V_i}$ 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 I(0) for the full sample over which the H-P filter is applied. However, due to the endpoint problem with H-P filters, this does not guarantee that the I(0) property will prevail when cointegration tests are applied since samples for individual vintages vary. The H-P filter was applied to the full available sample for each vintage.[19]

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 differences between the 2001 and 2004 vintages. Notice, too, that the longer the temporal difference between X_t^f and the particular $X_t^{V_i}$ 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

[18] 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.

[19] The data were also generated for cases where the H-P filter was applied to X_t^f for the same sample as that available for each X_t^v 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^v and the 2004Q1 vintage to emphasize the role of these long-run factors in influencing the time series properties of benchmark revisions.

($X_t^{2004Q1} - X_t^{2001Q1}$) is found to be I(1).[20] The principal 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.

As a preliminary to the estimates and forecasts using Phillips curves to come, we consider whether benchmark revisions, and their behavior toward the end of the sample, might reflect some underlying economic phenomenon that is only captured in subsequent data 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. The results that follow focus less on this possibility. Table 2 gives estimates of the simple correlation between pairs of revisions, again for our estimates of the output gap.[21] Taking the most recent revision as the benchmark (i.e., $R_t^{2001Q1} = X_\tau^{2004Q1} - X_\tau^{2001Q1}$), 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, $\rho(R_t^{2001Q1}, R_t^{1976Q1})$ is only 0.19 while $\rho(R_t^{1976Q1}, R_t^{1971Q1})$ 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

[20] 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).

[21] 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, generally are found to be I(1).

positive. Hence, differences between vintages at the time of benchmark revisions may have some predictive power for subsequent revisions.[22]

Table 2 also displays correlations between equity and housing returns, and benchmark revisions.[23] 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.[24]

3.3 Cointegration and Cointegration Breakdown Tests

The unit root tests shown in Table 1 assume a particular form for the cointegrating relationship, namely not restricting θ_0 in (5) but restricting θ_1 to be equal to 1. We may instead wish to formally estimate and test the joint restriction that $\theta_0 = 0$, $\theta_1 = 1$, in other words 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.

---

[22] 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.

[23] 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 US) available at www.ofheo.gov/download/asp. Log differencing was also applied to this time series.

[24] 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. See, however, van Norden (2005).

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 H-P 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 portion shows the estimated test statistics against the calculated critical values for the null hypothesis of stability at the end of each sample. These are given by the P_c test statistic (PC) while the calculated critical values are shown as P_{CV}.[25] The vertical lines indicate the year when a benchmark revision took place (i.e., every five years beginning in 1966 and ending in 2001). Test statistics and their critical values are shown for the last three years prior to a benchmark revision.

The bottom portion of Figure 3, plots the R_c test statistic against the calculated critical value for the null of a unit root disturbance in the second sub-sample. The vertical lines and the period over which the test statistics are shown in the top portion of Figure 3. Rejections of the

[25] The previous section briefly describes how the critical values are found. Detailed definitions and formulas for the test statistics can be found in Andrews and Kim (2003).

null are indicated by a test statistic that falls below the critical value.[26] All critical values are evaluated at a p-value equal to .05.

The top portion of Figure 3 clearly reveals that instability is a feature of the end of each of the samples considered as revealed by the test statistics that fall below the critical values in the last few observations. The bottom portion of Figure 3 reveals that, in most cases, the unit root null cannot be rejected in the second sub-sample. There are several non-rejections toward the end of the samples for the 1966, 1971, and 1976 benchmark revisions. In general, the results confirm 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

$$\pi_t^{for} = \alpha_0 + \alpha_1 \pi_{t-1}^{f} + \alpha_2 u_{t-1}^{v} + \varepsilon_t^{v} \qquad (9)$$

where π_t^{for} is the inflation forecast released by a public or private sector agency (see section 3.1), π_{t-1} is the actual inflation rate lagged and u_{t-1} is the error correction term from equation (5), that is, the one period lagged difference between a particular vintage of real GDP and the

[26] In other words, a test statistic that is below the critical value is equivalent to a p-value below .05.

final revision.[27] Equation (9) assumes that the current 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 in whether α_2 is significantly different from zero.[28] If benchmark revisions contain information that could have improved inflation forecasts then we would expect rejection of the null 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:

$$\pi_t^f = \beta_0 + \beta_1 \tilde{y}_{t-1}^v + \beta_2 \pi_{t-1}^f + \beta_3 E(\pi_{t+1}^f \left| \pi_{t-1}^f \right.) + \beta_4 u_{t-1}^v + \xi_t^v \qquad (10)$$

where $\tilde{y}^v$ is an estimate of the output gap, $E(\pi_{t+1}|\pi_{t-1})$ is the conditional expectation of next period's inflation rate (π_{t+1}) on lagged inflation (π_{t-1}), and all other terms have been previously 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 equation (10), two other variants are also 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

---

[27] 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 exists once a break is permitted. Some experimentation, however, reveals that the overall conclusions are unaffected by the manner in which u_{t-1} enters although some of the reported results are affected by this consideration (results not shown).

[28] The u_t in equation (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.

version of equation (10) with $\beta_2 = 0$, so that the resulting specification is entirely 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 (e.g., see James Stock's "Weak Instruments Web Page"; http://ksghome.harvard.edu/~JStock/aus/websupp/index.htm). In what follows, first, the usual, but not always ideal, strategy of using lagged values of $\tilde{y}_t$ and π_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 also relying on Stock's F-Test for instrument adequacy, and Andrews' GMM information criterion 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 the econometric issues.

Table 4 represents 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 the estimates of the various Phillips curves 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. If, however, the 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 GMMIC 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.

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 US inflation for the period 1960 to 2004. The three vertical long dashed lines identify three benchmark years, namely 1981, 1991, 1996, and 2001, that 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.[29] The gap between the two vertical dashed lines highlights the out-of-sample period over which inflation forecasts are generated. We report, in Table 5, both the one-step ahead forecast, together with its standard error, as well as the RMSE for one-year ahead forecasts (i.e., four steps ahead). 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 US economy was in a recession. The vertical shaded areas highlight recessionary periods as measured by NBER business cycle reference dates (http://www.nber.org/cycles.html/). Inflation

[29] More precisely, only version FL1 among the forward looking Phillips curve reported in Table 4 is considered.

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.

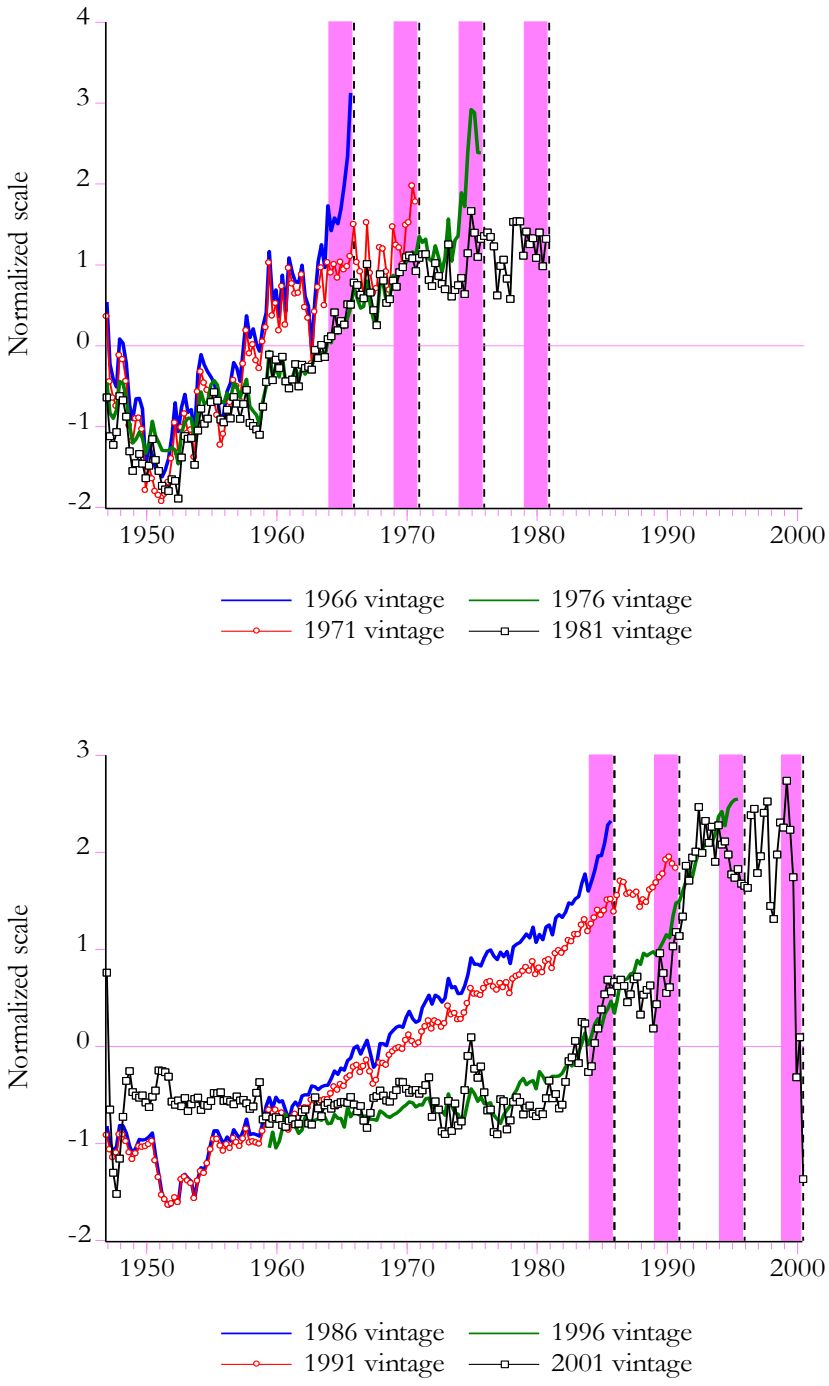
Part (b) of Figure 4 plots one through four steps ahead forecasts of inflation for the various cases considered. Forward-looking Phillips curve FL1 shown in Table 4 is estimated with or without the inclusion of benchmark revisions for samples that always begin in 1960Q1 and end in 1994Q4 in the case of the 1996 vintage. For the 1981 vintage the estimation period ends in 1979Q4, for the 1991 vintage the estimation period ends in 1989Q4, while 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 US 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.

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 differences in benchmark revisions are not cointegrated, and that the breakdown of the cointegration property seems to occur toward the end of the sample. Hence, we conclude that there is potentially some information content in benchmark revisions. In particular, we report estimates of Phillips curves which are augmented with benchmark revisions treated as an error correction term and find that these could have improved forecasts of inflation. A number of extensions and other aspects of the analysis in this paper remain to be dealt with. 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 may 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, rebasing, or some other factor can explain the potential role of revisions in forecasting inflation. Lastly, the results indicate that monetary policy is a ‘data rich’ environment (Bernanke and Boivin 2003) might also consider a role for the impact of revisions. These, and other, extensions are left for future research.

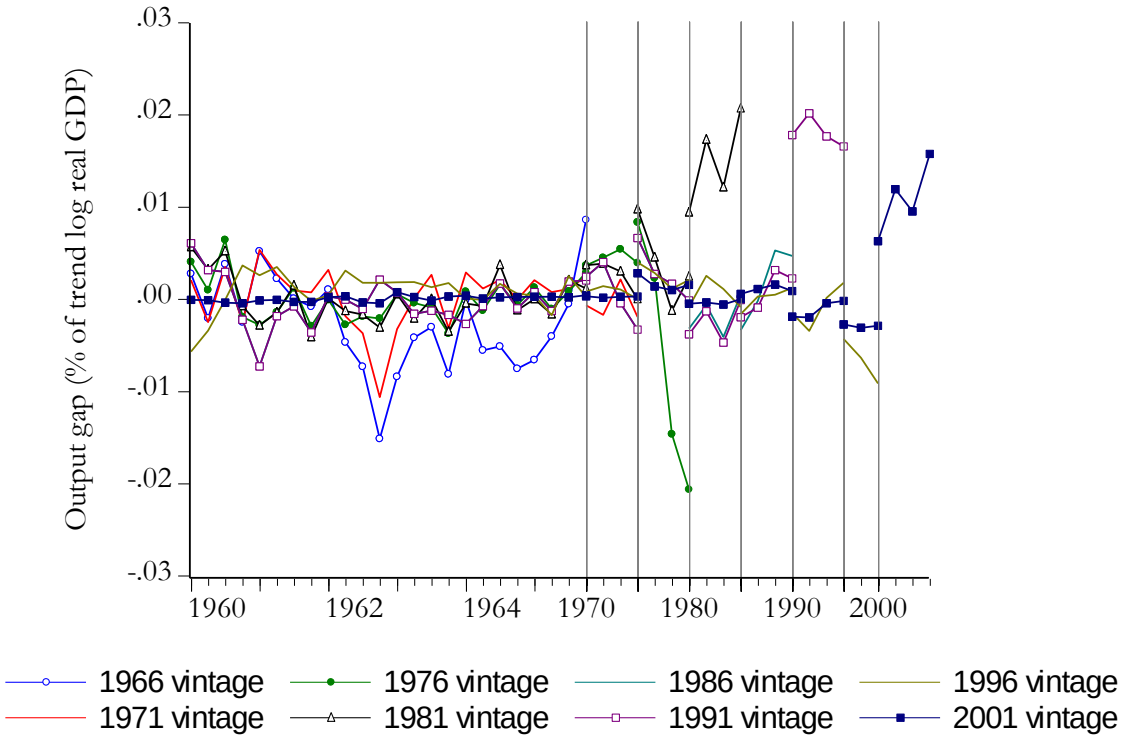
Figure 1. Stylized Facts About Benchmark Revisions

All Vintages and Observations (Top); All Vintages, Last Two Years of Data (Bottom)



Source: see text for data sources and series description.

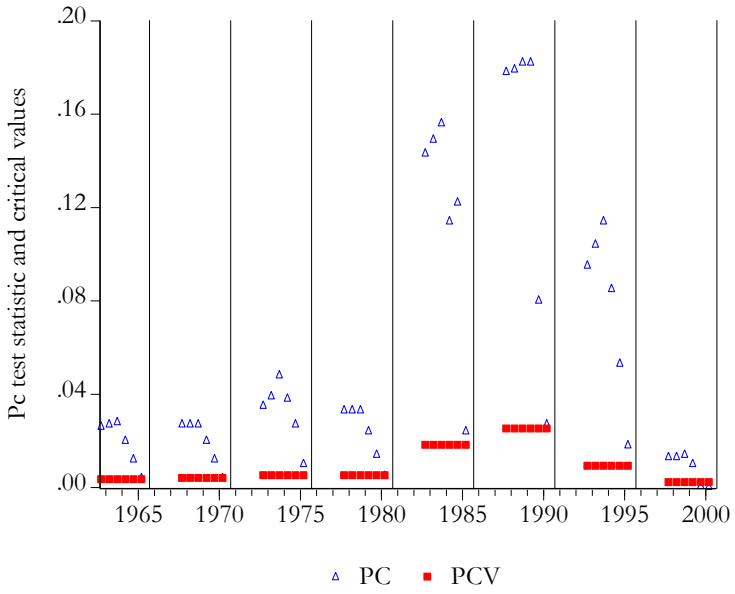
Figure 2. Benchmark Revisions in the Output Gap



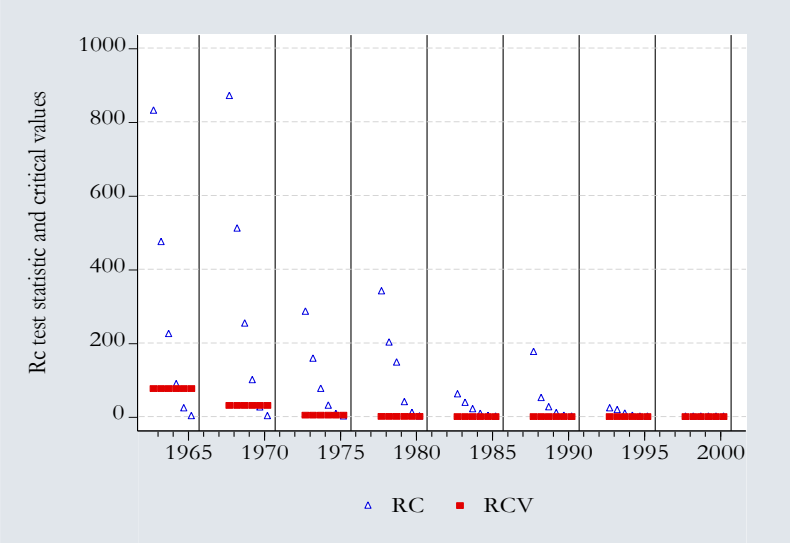
Sources: See text. Last two years of data only shown for all vintages after 1970. Also, see Figure 1. The output gap is found via H-P filtering (see text for estimation details) applied to the full available sample to each series individually.

Figure 3. End of Sample Cointegration Breakdown Tests

(a) Null hypothesis of end of sample stability

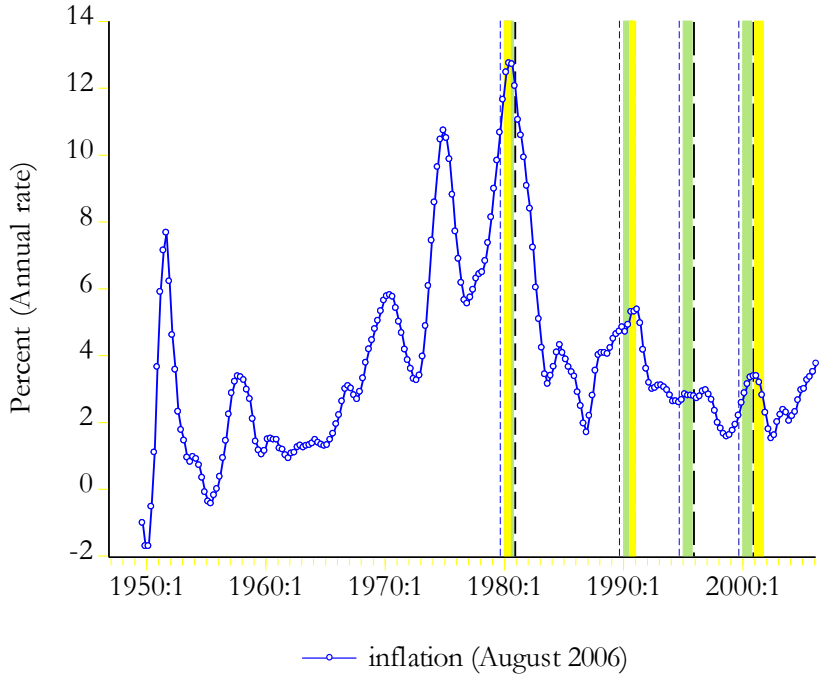


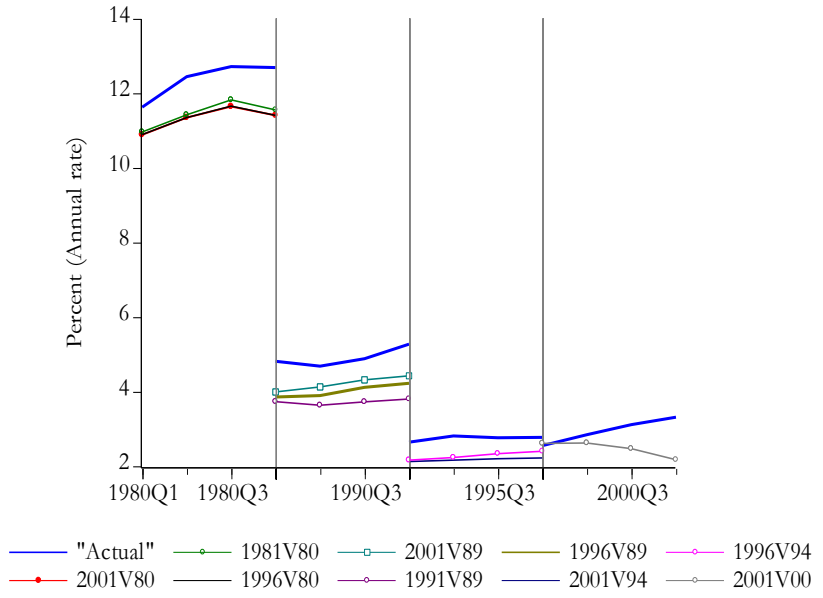
(b) Null hypothesis of a unit root disturbance



Note: Part (a) of the figure plots the test statistic (P_c) and the critical value P_{cv}; 5% level of significance) for the null that $\theta_0 = \theta, \forall_t = T \neq 1,\ldots,T+m$ where the test is carried out over the last two years of each sample. Also, see equation (8). Part (b) of the figure plots the test statistic R_c and R_{cv} for the null that $u_t \sim$ I(1), again over the last two years of each sample. See equation (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. US Inflation and Out-of-Sample Inflation Forecasts: The Role of Benchmark Revisions





Note: The 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 Roots in Benchmark Revisions

	Benchmark Revision							
Series	1966	1971	1976	1981	1986	1991	1996	2001
Real Cons.	1.13(4) [.98, 71]	0.42(4) [.98, 91]	0.01(.96) [.96, 115]	-0.18(0) [.94, 135]	-0.40(2) [.90, 153]	-0.77(2) [.83, 173]	-1.95(1) [.31, 143]	-0.34(2) [.92, 213]
Real GDP	0.70(1) [.99, 74]	0.16(3) [97, 92]	0.92(4) [.99, 111]	-0.38(5) [.91, 130]	1.44(0) [.99, 155]	0.62(0) [.99, 175]	3.70(2) 1.00, 142]	-1.46(6) [.55, 209]
Output Gap	**-4.55(0)** **[.00, 75]**	**-6.98(0)** **[.00, 95]**	**-5.20(0)** **[.00, 115]**	**-3.07(4)** **[.03, 131]**	**-5.84(0)** **[.00, 155]**	**-4.96(0)** **[.00, 175]**	**-5.13(0)** **[.00, 144]**	-0.67(4) [.85, 211]
Δln R.Cons	**-6.01(2)** **[.00, 69]**	**-3.35(4)** **[.02, 87]**	**-4.51(7)** **[.00, 104]**	-2.46(4) [.13, 127]	-2.13(8) [.23, 143]	**-5.47(4)** **[.00, 167]**	-1.61(4) [.48, 136]	**-5.71(11)** **[.00, 200]**
Δln R.GDP	**-2.94(5)** **[.05, 66]**	**-3.19(5)** **[.02, 86]**	**-4.07(1)** **[.00, 110]**	**-4.40(0)** **[.00, 131]**	**-3.83(4)** **[.00, 147]**	**-4.46(4)** **[.00, 167]**	**-4.53(0)** **[.00, 140]**	**-3.66(4)** **[.01, 207]**

Note: All tests are based on the Augmented Dickey-Fuller test statistic with the augmentation chosen according to the Akaike Information Criterion. No deterministic trends allowed in any test equation. In parenthesis, the lag length in the augmentation is shown. In brackets, the p-value followed by the number of observations after differencing. 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 equation (2), where X_t^f is the 2004 vintage, and X_t^v are for v=1966, …, 2001. Bold items signify rejections of the null of a unit root in R_t at conventional levels of significance.

Table 2. Simple Correlations Between Benchmark Revisions: Output Gap and Asset Prices

	Benchmark Revisions								Asset Price		
Benchmark Revisions	(1) 1966	(2) 1971	(3) 1976	(4) 1981	(5) 1986	(6) 1991	(7) 1996	(8) Bench-mark Revision year	(9) Equities S	(10) Housing H	(11) Fed funds ΔFFR
1966		0.97	0.68	0.68	0.23	0.23	0.32	1966	0.02	-	-0.49
1971			0.70	0.69	0.22	0.22	0.41	1971	-0.20	-	-0.47
1976				0.91	0.57	0.57	0.46	1976	-0.53	-	-0.46
1981					0.66	0.66	0.64	1981	-0.33	-0.02	-0.30
1986						0.98	0.43	1986	-0.05	-0.39	-0.16
1991							0.31	1991	0.00	-0.34	-0.16
1996								1996	0.02	-0.12	-0.25
2001	-0.06	-0.04	0.19	0.11	0.07	0.05	0.50	2001	-0.93	-0.15	-0.15

Notes: The values in the numbered columns show the simple correlation between pairs of benchmark revisions (e.g., the correlation between R_t^{1966} and R_t^{1981} is 0.68). The final three columns give 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_t^{1991} and H is -0.34).

Table 3. Cointegration Tests: Pairs of Vintages From 1966 to 2004

	Log real GDP: 2004 Vintage (1)		Log real consumption: 2004 Vintage (2)		Output Gap: 2004 Vintage (3)	
2001 vintage	(2) r=0	5.86	(1) r=0	4.58	(2) r=0	49.34*
	r=1	2.03	r=1	0.02	r=1	0.35*
1996 vintage	(2) r=0	14.10	(1) r=0	26.97*	(1) r=0	20.20*
	r=1	1.34	r=1	1.34	r=1	12.33*
1991 vintage	(2) r=0	12.13	(1) r=0	3.33	(2) r=0	44.23*
	r=1	0.18	r=1	1.89	r=1	20.97*
1986 vintage	(2) r=0	9.29	(1) r=0	2.93	(2) r=0	42.07*
	r=1	0.12	r=1	0.05	r=1	29.62*
1981 vintage	(2) r=0	12.72	(1) r=0	11.26	(2) r=0	35.47*
	r=1	0.87	r=1	2.87	r=1	10.50*
1976 vintage	(2) r=0	14.26	(1) r=0	5.65	(2) r=0	35.98*
	r=1	3.84	r=1	0.07	r=1	18.55*
1971 vintage	(2) r=0	12.85	(1) r=0	4.61	(2) r=0	34.67*
	r=1	0.27	r=1	0.05	r=1	24.50*
1966 vintage	(2) r=0	9.77	(1) r=0	5.65	(2) r=0	24.48*
	r=1	0.75	r=1	1.56	r=1	14.10*

Note: The test for cointegration is the Johansen VAR-based test with lag lengths, shown in parenthesis, chosen according to the Schwarz information criterion. The series are assumed to contain a constant and a linear trend, and the form of the cointegrating equation is as in equation (5). r=0, r=1 are the maximal eigenvalue test statistics for the null that there is no cointegration or, at most, one cointegrating vector, respectively, at the 5% level of significance.

Table 4. Inflation Forecasts and the Significance of Benchmark Revisions

	Dependent Variable: π_t^f						Dependent Variable: π_t					
	CONS		ECON		OECD		BL	BL	FL1	FL1	FL2	FL2
	1996	2001	1996	2001	1996	2001	1996	2001	1996	2001	1996	2001
Constant	2.20 (1.28)*	1.81 (.16)*	.92 (.30)*	.87 (.17)*	.88 (.16)*	.68 (.14)*	.04 (.05)	.03 (.05)	-.05 (.02)*	-.04 (.02)*	-.05 (.02)*	-.02 (.02)*
π_{t-1}	.47 (.09)*	.56 (.06)*	.89 (.10)*	.90 (.07)*	.78 (.03)*	.80 (.03)*	.99 (.01)*	.99 (.01)*	.45 (.02)*	.46 (.02)*	.47 (.02)*	.46 (.02)*
$E(\pi_{t+1}\|\pi_{T-1})$	-	-	-	-	-	-	-	-	.56 (.03)*	.56 (.02)*	.55 (.02)*	.55 (.02)
$\tilde{y}_{t-1}$	-	-	-	-	-	-	.19 (.02)*	.19 (.16)*	-.05 (.01)*	-.04 (.01)*	-.01 (.01)	-.03 (.01)*
u_{t-1}	-.37 (.24)	-.32 (.11)*	-.53 (.25)**	-.01 (.11)	-.80 (.42)$^+$	-.44 (.33)	.31 (.14)**	.08 (.10)	-.09 (.08)	-.50 (0.2)$^+$	-	-
Summary Statistics												
$\overline{R}^2$	.52	.73	.79	.82	.82	.83	.98	.99	.99	.99	.99	.99
Sample	1989:4-1995:4	1989:4-2001:1	1991:3-1995:4	1990:3-2001:1	1960:3-1995:4	1960:3-2000:4	1959:4-1995:4	1959:4-2001:1	1960:3-1995:4	1960:3-2001:1	1960:3-1995:4	1960:3-2001:1
J	-	-	-	-	-	-	-	-	7.1 (.79)	8.15(.85)	9.94(.87)	6.52(.63)
GMMIC	-	-	-	-	-	-	-	-	-24.73	-25.42	-29.66	-30.52
F	-	-	-	-	-	-	-	-	718.1	891.9	793.2	629.4

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 means the backward-looking Phillips curve is estimated; FL is a 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, GMMIC is the GMM Information Criterion due to Andrews (1999) and F is the Stock's F-statistics for instrument adequacy. (see Stock, Wright, and Yogo (2002)). * signifies statistically significant at the 1% level, ** at the 5% level, and + at the 10% level.

Table 5. Out of Sample Inflation Forecasts and the Significance of Benchmark Revisions

Vintage	Estimation Period	RMSE With u_{t-1}	RMSE Without u_{t-1}	One step ahead forecast (s.e.) With u_{t-1}	One step ahead forecast (s.e.) Without u_{t-1}	Actual inflation
1996	1960.1-1994.4	0.47	0.58	3.10% (.33)	2.16 (.13)	2.64%
1996	1960.1-1989.4	0.90	1.11	4.27% (.31)	3.85 (.13)	4.81%
1996	1960.1-1979.4	1.06	1.61	10.89% (.11)	10.88 (.11)	11.62%
2001	1960.1-1999.4	0.66	0.78	3.85% (.12)	3.73 (.13)	2.55%
2001	1960.1-1994.4	0.57	0.68	3.26% (.24)	2.12 (.12)	2.64%
2001	1960.1-1989.4	0.71	1.16	3.99% (.09)	4.21 (.25)	4.81%
2001	1960.1-1979.4	1.07	1.61	10.88% (.12)	10.76 (.13)	11.62%

Note: RMSE= root mean squared 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 root mean squared errors of those forecasts are reported in the Table. The one steap ahead forecast is for the quarter that immediately follows the end of the estimation sample.

References

Andrews, D. (2003), "End-of-Sample Instability Tests", *Econometrica* 71(6): 1661-1694.

Andrews, D. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica* 61(4): 821-56.

Andrews, D., and J.-Y. Kim (2003), "End-of-Sample Cointegration Breakdown Tests", Cowles Foundation working paper 1404.

Bernanke, B.S., and J. Boivin (2003), "Monetary Policy in a Data-Rich Environment", *Journal of Monetary Economics* 50 (April): 525-46.

Croushore, D., and C. L. Evans (2006), "Data Revisions and the Identification of Monetary Policy Shocks", *Journal of Monetary Economics* (forthcoming).

Coushore, D., and T. Stark (2005), "A Real Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics* 85 (August): 605-17.

Coushore, D., and T. Stark (2002), "A Real Time Data Set for Macroeconomists", *Journal of Econometrics* 105: 111-130.

D'Agostino, A., D. Giannone, and P. Surico (2006), "(Un)Predictability and Macroeconomic Stability", ECB working paper No. 605, April.

Faust, J., J. Rogers, and J. Wright (2005), "News and Noise in G7 Announcements", *Journal of Money, Credit and Banking* 37 (June): 403-19.

Garratt, A., and S. Vahey (2006), "UK Real-Time Macro Data Characteristics", *Economic Journal* 116 (February): F119-F135.

Golinelli, R., and G. Pangi (2005), "Short-Run Station GDP Forecasting and Real Time Data", GDPR Discussion Paper 5302, October.

Gregory, A., and B. Hansen (1996), "Tests for Cointegration in Models with Regime and Trend Shifts", *Oxford Bulletin of Economics and Statistics* 58 (August): 555-60.

Hall, A. (2005), *Generalized Method of Moments* (Oxford: Oxford University Press).

Hansen, H., and S. Johansen (1999), "Some Tests for Parameter Constancy in Cointegrating VAR Models", Econometrics Journal 2(2): 306-33.

Hermann, H., A. Orphanides, and P.L. Siklos (2005), "Real-Time Data and Monetary Policy", *North American Journal of Economics and Finance* 16 (December): 271-76.

Kozicki, S. (2004), "How Do Data Revisions Affect the Evaluation and Conduct of Monetary Policy?" *Economic Review*, Federal Reserve Bank of Kansas City, Quarter 1, pp. 5-38.

Mankiw, N.G., and M. Shapiro (1986), "News or Noise: An Analysis of GNP Revisions", *Survey of Current Business* (May): 20-25.

Mankiw, N.G., D. Runkle, and M. Shapiro (1984), "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics* 14 (July): 15-27.

Meyer, L. (2004), *A Term at the Fed: An Insider's View* (New York: Harpers Collins).

Orphanides, A., and S. van Norden (2005), "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time", *Journal of Money, Credit and Banking*, 37 (June): 583-601.

Orphanides, A. (2001), "Monetary Policy Rules Based on Real-Time Data", *American Economic Review* 91 (September): 964-85.

Ritter, J.A. (2000), "Feeding the National Accounts", *Review* of the Federal Reserve Bank of St. Louis 82 (March/April): 11-20.

Runkle, D. (1998), "Revisionist History: How Data Revisions Distort Economic Policy Research", *Quarterly Review*, Federal Reserve Bank of Minneapolis 22 (Fall): 3-12.

Siklos, P.L., and M. Wohar (2006), "Estimating Taylor-Type Rules: An Unbalanced Regression?" in T. Fomby and D. Terrell (Eds.), *Advances in Econometrics*, vol. 2 (Amsterdam: North-Holland).

Siklos, P.L. (2002), *The Changing Face of Central Banking: Evolutionary Trends Since World War II* (Cambridge: Cambridge University Press).

Siklos, P.L. (1996), "An Empirical Analysis of Revisions in U.S. Macroeconomic Aggregates", *Applied Financial Economics* 6 (February): 59-70.

Siklos, P.L., and C.W.J. Granger (1997), "Regime-Sensitive Cointegration with An Illustration from Interest Rate Parity", *Macroeconomic Dynamics* 1 (3, 1997): 640-57.

Stock, J. H., J. H. Wright, and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Methods of Moments", *Journal of Business and Economic Statistics* 20, 518 - 529.

Van Norden, S. (2005), "Are We There Yet? Looking for Evidence of a New Economy", working paper, GEC Montreal, CIRANO & CIREQ.